

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*

Abstract—In compressed sensing (CS) MRI, model-based methods are pivotal to achieving accurate reconstruction. One of the main challenges in model-based methods is finding an effective prior to describe the statistical distribution of the target image. Plug-and-Play (PnP) and REGularization by Denoising (RED) are two general frameworks that use denoisers as the prior. While PnP/RED methods with convolutional neural network (CNN) based denoisers outperform classical hand-crafted priors in CS MRI, their convergence theory relies on assumptions that do not hold for practical CNN models. The recently developed gradient-driven denoisers offer a framework that bridges the gap between practical performance and theoretical guarantees. However, the numerical solvers for the associated minimization problem remain slow for CS MRI reconstruction. This paper proposes a complex quasi-Newton proximal method that achieves faster convergence than existing approaches. To address the complex domain in CS MRI, we propose a modified Hessian estimation method that guarantees Hermitian positive definiteness. Furthermore, we provide a rigorous convergence analysis of the proposed method for nonconvex settings. Numerical experiments on both Cartesian and non-Cartesian sampling trajectories demonstrate the effectiveness and efficiency of our approach.

Index Terms—CS MRI, gradient-driven denoiser, second-order, convergence, complex domain, spiral and radial acquisitions.

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is a non-invasive imaging technique that generates images of the internal structures of the body [1]. MRI is widely used in clinical settings for disease diagnosis, treatment guidance, and functional and advanced imaging, among other applications [2]. In practice, MRI scanners acquire k-space data, which are the

Fourier components of the desired images. The acquisition procedure is slow, leading to patient discomfort, increased motion artifacts, reduced clinical efficiency, and other issues. To accelerate the acquisition, modern MRI scanners use multiple coils (parallel imaging) to acquire less Fourier components. The parallel imaging technique incorporates additional spatial information that can help significantly reduce MRI acquisition time [3, 4]. Moreover, by combining with compressed sensing (CS) [5], one can acquire even fewer Fourier components, further accelerating the acquisition process. However, the reconstruction in CS MRI requires iterative solvers for the following composite minimization problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{C}^N} F(\mathbf{x}) \equiv \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{h(\mathbf{x})} + \lambda f(\mathbf{x}), \quad (1)$$

where $\mathbf{A} \in \mathbb{C}^{MC \times N}$ denotes the forward model specifying the mapping from the image $\mathbf{x} \in \mathbb{C}^N$ to the k-space data $\mathbf{y} \in \mathbb{C}^{ML}$, and $f(\mathbf{x})$ refers to the regularizer describing the prior information about $\mathbf{x}$. In practice, we have $M \ll N$ due to downsampling. The trade-off parameter $\lambda > 0$ balances the data-fidelity term $h(\mathbf{x})$ and the regularizer $f(\mathbf{x})$. Here $C > 1$ denotes the number of coils. The system matrix $\mathbf{A}$ is a stack of submatrices $\mathbf{A}_c$ such that $\mathbf{A} = [\mathbf{A}_1; \mathbf{A}_2; \dots; \mathbf{A}_C]$. The submatrices $\mathbf{A}_c$ are defined as $\mathbf{A}_c \in \mathbb{C}^{M \times N} = \mathbf{P}\mathbf{F}\mathbf{S}_c$ where $\mathbf{P}$ is the downsampling pattern, $\mathbf{F}$ defines the (non-uniform) Fourier transform, and $\mathbf{S}_c$ denotes the coil sensitivity map for c th coil, which is patient specific.

The data fidelity term $h(\mathbf{x})$ enhances the data consistency. Since the k-space data is highly downsampled, the regularizer $f(\mathbf{x})$ is required to stabilize the solution. The choice of f can significantly affect the reconstruction quality. Classical hand-crafted regularizers have proven effective for MRI reconstruction including wavelets [6, 7], total variation (TV) [8], a combination of wavelet and TV [5, 9], dictionary learning [10, 11], and low-rank methods [12], to name a few. See [13, 14] for a review of various choices for f .

Over the past decade, deep learning (DL) has attracted significant attention for MRI reconstruction because of its superior performance [15]. Unlike hand-crafted regularizers, DL learns complex image priors directly from large amounts of data. The promising DL-based methods for MRI reconstruction include end-to-end learning frameworks [16] and physics-informed deep unrolling methods [17–21]. More recently, generative models have emerged as powerful tools

T. Hong is with the Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA (Email: tao.hong@austin.utexas.edu). TH was with the Department of Radiology, University of Michigan, Ann Arbor, MI 48109, USA. TH was partially supported by NIH grant R01NS112233.

Z. Xu is with the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA (Email: zhaoyix@umich.edu).

S. Chun is with the Department of ECE, INMC & IPAI, SNU, Seoul, 08826, South Korea (Email: sychun@snu.ac.kr). SC was partially supported by IITP grant [No.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] and NRF grant (No. RS-2025-02263628), all funded by the Korea government(MSIT).

L. Hernandez-Garcia is with the Department of Radiology, University of Michigan, Ann Arbor, MI 48109, USA (Email: hernan@umich.edu). LH was partially supported by NIH grant R01NS112233.

J. Fessler is with the EECS Department, University of Michigan, Ann Arbor, MI 48109, USA (Email: fessler@umich.edu). Supported in part by NIH grants R01EB035618 and R21EB034344.

for learning priors in MRI reconstruction, gaining substantial interest [22, 23].

An alternative framework to DL is Plug-and-Play (PnP)/REgularization by Denoising (RED) [24, 25], which leverages the most effective denoisers, such as BM3D [26] or DnCNN [27], achieving outstanding performance in various imaging tasks [28–33]. Compared to the end-to-end and unrolling DL approaches, which are typically designed for a predefined imaging task and rely on training with massive amounts of data, PnP/RED can be easily adapted to specific applications without requiring retraining. This capability is especially advantageous for addressing CS MRI problems, where sampling patterns, coil sensitivity maps, and image resolutions can vary greatly between scans. Detailed discussions about using PnP for MRI reconstruction are found in [34]. The following subsections first introduce background on PnP/RED priors and related theoretical work. We then discuss the gradient-driven denoisers framework and the associated minimization problem.

A. Inverse Problems with PnP/RED Priors

Proximal algorithms [35] are a class of iterative methods for solving (1). At k th iteration, the proximal gradient method (PGM) is expressed as

$$\mathbf{x}_{k+1} = \text{prox}_{\alpha\lambda f}^{\mathbf{I}}(\mathbf{x}_k - \alpha\nabla h(\mathbf{x}_k)), \quad (2)$$

where $\alpha \in \mathbb{R}$, $\alpha > 0$ is the stepsize, $\nabla h(\cdot)$ denotes the gradient of $h(\cdot)$, and $\text{prox}_{\alpha\lambda f}^{\mathbf{W}}(\cdot)$ denotes the weighted proximal mapping (WPM) defined as

$$\text{prox}_{\alpha\lambda f}^{\mathbf{W}}(\cdot) \triangleq \arg \min_{\mathbf{x} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{x} - \cdot\|_{\mathbf{W}}^2 + \alpha\lambda f(\mathbf{x}), \quad (3)$$

where $\mathbf{W} \in \mathbb{C}^{N \times N}$, $\mathbf{W} \succ 0$ is a Hermitian positive definite matrix and $\|\mathbf{x}\|_{\mathbf{W}}^2 = \mathbf{x}^H \mathbf{W} \mathbf{x}$. Here, $\mathbf{H}$ denotes the Hermitian transpose operator. Clearly, $\text{prox}_{\alpha\lambda f}^{\mathbf{I}}(\cdot)$ represents the classical proximal operator [35]. Since $\text{prox}_{\alpha\lambda f}^{\mathbf{I}}(\cdot)$ can be interpreted as a denoiser, the PnP-ISTA algorithm simply replaces the proximal operator with an arbitrary denoiser $\mathbf{D}(\cdot)$. In practice, the sequence $\{\mathbf{x}_k\}$ generated by PnP-ISTA generally converges slowly. To address this issue, Pendu et al. [36] introduced a preconditioned PnP-ADMM algorithm that uses a diagonal matrix as the preconditioner. More recently, Hong et al. [37] presented a provably convergent preconditioned PnP framework that supports general preconditioners and demonstrates fast convergence in CS MRI reconstruction.

An alternative to PnP, RED [25, 38, 39] introduces an explicit regularization term based on a denoiser, i.e., $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T [\mathbf{x} - \mathbf{D}(\mathbf{x})]$. Here, $\mathbf{T}$ denotes the transpose operator. Romano et al. [25] demonstrated that if the denoiser satisfies the local homogeneity property, the gradient of $f(\mathbf{x})$ in RED can be expressed as $\mathbf{x} - \mathbf{D}(\mathbf{x})$. Since $f(\mathbf{x})$ in RED is differentiable, various iterative methods, such as gradient descent, proximal methods, and quasi-Newton methods, can be applied to solve RED, provided that $h(\mathbf{x})$ is also differentiable. To accelerate convergence in RED, several techniques have been adopted, such as vector extrapolation [38], fast proximal methods [39], and weighted proximal methods [40], among others.

While PnP/RED has demonstrated significant empirical success, theoretical research on its convergence continues to be an active area of study, see [25, 31, 39, 41–50]. These studies assume either that the denoiser approximates the MAP or MMSE estimator, that $\|\mathbf{D}(\mathbf{x}) - \mathbf{x}\|$ is upper bounded, or that it is nonexpansive, satisfying

$$\|\mathbf{D}(\mathbf{x}_1) - \mathbf{D}(\mathbf{x}_2)\| \leq \|\mathbf{x}_1 - \mathbf{x}_2\|. \quad (4)$$

However, it is known that if the Jacobian of such estimators is well-defined, it must be symmetric [39, 51, 52]. However, RED/PnP frequently achieves state-of-the-art performance with denoisers that do not satisfy this condition, such as Convolutional Neural Network (CNN) based denoisers. So RED/PnP cannot be explained as an optimization-based solver. Although optimization-free explanations for RED and PnP exist, understanding their behaviors and characterizing their solutions remains challenging. Moreover, most commonly used denoisers do not satisfy (4). Another approach [43, 53, 54] is to train a denoiser by incorporating a regularization term in training to constrain its Lipschitz constant. However, this approach remains an open challenge, as there is currently no practical way to strictly guarantee that the trained denoiser is nonexpansive [55, 56].

B. Inverse Problems with gradient-driven Denoisers

To address the aforementioned challenges and bridge the gap between theoretical assumptions and practical performance in PnP/RED, recent works [57–60] have proposed constructing gradient-driven denoisers. In this framework, the denoiser is given by the difference between the noisy image $\mathbf{x}$ and the gradient of a scalar-valued function $f_{\theta}(\mathbf{x})$, i.e.,

$$\mathbf{D}_{\theta}(\mathbf{x}) \equiv \mathbf{x} - \nabla_{\mathbf{x}} f_{\theta}(\mathbf{x}). \quad (5)$$

In practice, $f_{\theta}(\mathbf{x})$ is constructed with a scalar-output CNN and trained so that $\mathbf{x} - \nabla_{\mathbf{x}} f_{\theta}(\mathbf{x})$ serves as a denoiser. Here θ denotes the trained network parameters. In medical imaging reconstruction, interpretability, reliability, and provable numerical algorithms are crucial, as the reconstructed images are commonly used for disease diagnosis. Therefore, it is essential to understand how the underlying algorithms function. The use of gradient-driven denoisers for CS MRI reconstruction enables the integration of deep learning techniques while maintaining interpretability and reliability. Specifically, we aim to solve the following optimization problem to recover the latent image:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{C}} F(\mathbf{x}) \equiv \underbrace{\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}_{h(\mathbf{x})} + \lambda f_{\theta}(\mathbf{x}), \quad (6)$$

where $\mathcal{C}$ is a closed convex set of $\mathbb{C}^N$. For notational simplicity, we ignore θ in $f_{\theta}(\mathbf{x})$ and simply write $f(\mathbf{x})$ hereafter. Similarly, we use $\nabla f(\mathbf{x})$ to denote $\nabla_{\mathbf{x}} f(\mathbf{x})$. Once the trade-off parameter λ is chosen, we fix it throughout the minimization. Therefore, we absorb λ into f hereafter and write $f(\mathbf{x})$ instead of $\lambda f(\mathbf{x})$.

Although both $f(\mathbf{x})$ and $h(\mathbf{x})$ are differentiable, $f(\mathbf{x})$ may be nonconvex. Therefore, the underlying numerical algorithms

for solving (6) should be theoretically sound and applicable to nonconvex settings. Cohen et al. [57] employed the projected gradient method with a line search strategy to solve (6) and established its convergence under the assumption that the gradient of $f(\mathbf{x})$ is Lipschitz continuous. Hurault et al. [58] applied the proximal gradient method, where at the k th iteration, $\mathbf{x}_{k+1}$ is updated as follows:

$$\mathbf{x}_{k+1} = \text{prox}_{\alpha_k h + \iota_{\mathcal{C}}}^{\mathbf{I}}(\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)), \quad (7)$$

where α_k represents the stepsize and $\iota_{\mathcal{C}}(\mathbf{x})$ denotes the characteristic function such that

$$\iota_{\mathcal{C}}(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in \mathcal{C}, \\ +\infty, & \text{otherwise.} \end{cases} \quad (8)$$

Similar to [57], the convergence is established under the same assumption and α_k is determined using a line search strategy. Note that both methods were tested on image deblurring and super-resolution tasks in the *real* domain. A direct extension of these methods to CS MRI reconstruction requires hundreds of iterations to converge, limiting their practical applicability. Therefore, improving the convergence speed is essential for CS MRI reconstruction with gradient-driven denoisers.

Alternatively, Tan et al. [61] adopted the method proposed in [62] to solve the following problem, instead of (6), by exploiting second-order information:

$$\min_{\mathbf{x} \in \mathbb{R}^N} h(\mathbf{x}) + \lambda \underbrace{\left(f(\mathbf{D}^{-1}(\mathbf{x})) - \frac{1}{2} \|\mathbf{D}^{-1}(\mathbf{x}) - \mathbf{x}\|_2^2 \right)}_{\tilde{f}(\mathbf{x})}, \quad (9)$$

where $\mathbf{D}(\mathbf{x}) = \mathbf{x} - \nabla f(\mathbf{x})$ denotes the denoiser. Their algorithmic routine, dubbed PnP-LBFGS, treats $\mathbf{D}(\mathbf{x})$ as a proximal mapping of $\lambda f(\mathbf{x})$. However, PnP-LBFGS requires the Lipschitz constant of ∇f to be upper bounded by one, which is challenging to guarantee in practice. Furthermore, Tan et al. only considered image deblurring and super-resolution problems for *real-valued* images, and it remains unclear how convex constraints could be incorporated into their framework.

C. Contributions and Roadmap

Motivated by recent work demonstrating the potential advantages of using second-order information for acceleration in many applications [9, 40, 63–70], we propose a *complex* quasi-Newton proximal method (CQNPM) that incorporates additional Hessian information at each iteration to accelerate the convergence of solving (7). In contrast to [9], we estimate the Hessian matrix of $f(\mathbf{x})$ rather than that of $h(\mathbf{x})$. Although $f(\mathbf{x})$ is differentiable, it is nonconvex and $\mathbf{x}$ is *complex*, which introduces new challenges in estimating a Hermitian positive definite Hessian matrix. To address these challenges, we develop an approach that enforces the estimated Hessian matrix to be Hermitian positive definite.

The main contributions of this paper are summarized as follows:

- We extend the gradient-driven denoisers to complex-valued CS MRI reconstruction. Moreover, we propose a *complex* quasi-Newton proximal approach to efficiently solve the associated minimization problem.

- We propose a modified Hessian estimation method that enforces a Hermitian positive definite Hessian matrix. Moreover, we provide a rigorous convergence analysis of the proposed approach under the nonconvex settings.
- We extensively validate the performance of our method using both Cartesian and non-Cartesian sampling trajectories on brain and knee images.

The rest of this paper is organized as follows. Section II introduces our proposed method and explains how to estimate a Hermitian positive definite Hessian matrix. In addition, Section III provides a rigorous convergence analysis. Finally, Section IV presents numerical experiments that study the performance of our method and validate the theoretical analysis. Supplementary material is also provided, including the validation of Assumption 1, a comparison with PnP-LBFGS, and additional experimental results.

II. PROPOSED METHOD

This section first introduces our approach for addressing (6). We then describe how to estimate a Hermitian positive definite Hessian matrix. Finally, we present an efficient strategy for solving the associated WPM.

Given an estimated Hessian matrix $\mathbf{B}_k$, at the k th iteration, we update the next iterate by solving a WPM, i.e.,

$$\mathbf{x}_{k+1} = \text{prox}_{\alpha_k h + \iota_{\mathcal{C}}}^{\mathbf{B}_k}(\mathbf{x}_k - \alpha_k \mathbf{H}_k \nabla f(\mathbf{x}_k)), \quad (10)$$

where $\mathbf{H}_k = \mathbf{B}_k^{-1}$. Unlike the usual PGM in (2) that uses the proximal mapping of the *regularizer*, the WPM in (10) uses the proximal mapping of the *data term* h plus the characteristic function in (8). Algorithm 1 summarizes the proposed method.

To compute a Hermitian positive definite approximation to the Hessian matrix, we introduce a modified memory-efficient self-scaling Hermitian rank-1 method (MMESHR1) that extends the method proposed in [37, 71–73]. Algorithm 2 presents the detailed steps of MMESHR1. The operator $\Re(\cdot)$ in Algorithm 2 denotes an operation that extracts the real part. Without the operator $\Re(\cdot)$, Algorithm 2 reduces to the Hessian approximation method proposed in [9]. For the problems in [9], we estimated the Hessian matrix for $h(\mathbf{x})$ so that $\langle \mathbf{m}_k, \mathbf{s}_k \rangle$ is guaranteed to be real for all iterations. However, $\langle \mathbf{m}_k, \mathbf{s}_k \rangle$ becomes complex in our problem setting, leading to a non-Hermitian Hessian approximation. Let $\bar{\mathbf{B}}_k$ denote the approximate Hessian matrix obtained using the approach proposed in [9]. To obtain a Hermitian positive definite Hessian approximation, we seek a $\mathbf{B}_k$ that is the nearest Hermitian approximation to $\bar{\mathbf{B}}_k$ in terms of Frobenius norm minimization, resulting in $\mathbf{B}_k = (\bar{\mathbf{B}}_k + \bar{\mathbf{B}}_k^H)/2$. In practice, this step is equivalent to applying $\Re(\cdot)$ at steps 2, 3, and 4 in Algorithm 2. Lemma 4 shows that the approximate Hessian matrices generated by Algorithm 2 are always Hermitian positive definite, even if f is nonconvex.

Computing the WPM in (10) is equivalent to solving the following minimization problem:

$$\arg \min_{\bar{\mathbf{x}} \in \mathcal{C}} G_k(\bar{\mathbf{x}}) \equiv \frac{1}{2} \left(\|\bar{\mathbf{x}} - \mathbf{w}_k\|_{\mathbf{B}_k}^2 + \alpha_k \|\mathbf{A}\bar{\mathbf{x}} - \mathbf{y}\|_2^2 \right), \quad (11)$$

Algorithm 1 Complex Quasi-Newton Proximal Method (CQNPM)

Initialization: $\mathbf{x}_1$ and stepsize $\alpha_k > 0$

Iteration:

- 1: **for** $k = 1, 2, \dots$ **do**
 - 2: Set $\mathbf{H}_k$ and $\mathbf{B}_k$ using Algorithm 2
 - 3: $\mathbf{x}_{k+1} \leftarrow \text{prox}_{\alpha_k \mathbf{H}_k + \iota_{\mathcal{C}}}(\mathbf{x}_k - \alpha_k \mathbf{H}_k \nabla f(\mathbf{x}_k))$
 - 4: **end for**
-

Algorithm 2 Modified Memory Efficient Self-Scaling Hermitian Rank-1 Method

Initialization: $\mathbf{x}_{k-1}$, $\mathbf{x}_k$, $\nabla f(\mathbf{x}_{k-1})$, $\nabla f(\mathbf{x}_k)$, $\delta = 10^{-8}$, $\theta_1 \in (0, 1)$, and $\theta_2 \in (1, \infty)$

- 1: Set $\mathbf{s}_k \leftarrow \mathbf{x}_k - \mathbf{x}_{k-1}$ and $\mathbf{m}_k \leftarrow (\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1}))$
- 2: Compute β such that

$$\min_{\beta} \{ \beta \in [0, 1] | \mathbf{v}_k = \beta \mathbf{s}_k + (1 - \beta) \mathbf{m}_k \} \text{ satisfies } \theta_1 \leq \frac{\Re(\langle \mathbf{s}_k, \mathbf{v}_k \rangle)}{\langle \mathbf{s}_k, \mathbf{s}_k \rangle} \text{ and } \frac{\langle \mathbf{v}_k, \mathbf{v}_k \rangle}{\Re(\langle \mathbf{s}_k, \mathbf{v}_k \rangle)} \leq \theta_2 \quad (13)$$

- 3: Compute $\tau_k \leftarrow \frac{\langle \mathbf{s}_k, \mathbf{s}_k \rangle}{\Re(\langle \mathbf{s}_k, \mathbf{v}_k \rangle)} - \sqrt{\left(\frac{\langle \mathbf{s}_k, \mathbf{s}_k \rangle}{\Re(\langle \mathbf{s}_k, \mathbf{v}_k \rangle)} \right)^2 - \frac{\langle \mathbf{s}_k, \mathbf{s}_k \rangle}{\langle \mathbf{v}_k, \mathbf{v}_k \rangle}}$
 - 4: $\rho_k \leftarrow \Re(\langle \mathbf{s}_k - \tau_k \mathbf{v}_k, \mathbf{v}_k \rangle)$
 - 5: **if** $\rho_k \leq \delta \|\mathbf{s}_k - \tau_k \mathbf{v}_k\| \|\mathbf{v}_k\|$ **then**
 - 6: $\mathbf{u}_k \leftarrow \mathbf{0}$
 - 7: **else**
 - 8: $\mathbf{u}_k \leftarrow \mathbf{s}_k - \tau_k \mathbf{v}_k$
 - 9: **end if**
 - 10: $\rho_k^{\mathbf{B}} \leftarrow \tau_k^2 \rho_k + \tau_k \mathbf{u}_k^{\mathbf{H}} \mathbf{u}_k$
 - 11: **Return:** $\mathbf{H}_k \leftarrow \tau_k \mathbf{I}_N + \frac{\mathbf{u}_k \mathbf{u}_k^{\mathbf{H}}}{\rho_k}$ and $\mathbf{B}_k \leftarrow \tau_k^{-1} \mathbf{I}_N - \frac{\mathbf{u}_k \mathbf{u}_k^{\mathbf{H}}}{\rho_k^{\mathbf{B}}}$
-

where $\mathbf{w}_k = \mathbf{x}_k - \alpha_k \mathbf{H}_k \nabla f(\mathbf{x}_k)$. Since the cost function in (11) is differentiable and strongly convex, we apply the accelerated projection method with fixed momentum to solve it [74, 75]. To avoid using a line search or computing the exact Lipschitz constant, which would increase the computational cost, we instead estimate an upper bound. The Hessian matrix of the cost function in (11) is $\mathbf{B}_k + \alpha_k \mathbf{A}^{\mathbf{H}} \mathbf{A}$. Clearly, we have the following relations

$$\begin{aligned} \mu_{\min}(\mathbf{B}_k + \mathbf{A}^{\mathbf{H}} \mathbf{A}) &\geq \mu_{\min}(\mathbf{B}_k) + \mu_{\min}(\mathbf{A}^{\mathbf{H}} \mathbf{A}), \\ \mu_{\max}(\mathbf{B}_k + \mathbf{A}^{\mathbf{H}} \mathbf{A}) &\leq \mu_{\max}(\mathbf{B}_k) + \mu_{\max}(\mathbf{A}^{\mathbf{H}} \mathbf{A}), \end{aligned} \quad (12)$$

where $\mu_{\min}(\cdot)$ and $\mu_{\max}(\cdot)$ denote the smallest and largest eigenvalues, respectively. Let $L_{\mathbf{A}}$ denote the largest eigenvalue of $\mathbf{A}^{\mathbf{H}} \mathbf{A}$. In CS MRI, the kspace data is under-sampled so that the smallest eigenvalue of $\mathbf{A}^{\mathbf{H}} \mathbf{A}$ is zero. Using the relations in (12) and $\mathbf{B}_k = \tau_k^{-1} \mathbf{I}_N - (\mathbf{u}_k \mathbf{u}_k^{\mathbf{H}})/(\rho_k^{\mathbf{B}})$ (cf. Algorithm 2), the largest and smallest eigenvalues of $\mathbf{B}_k + \alpha_k \mathbf{A}^{\mathbf{H}} \mathbf{A}$ are upper and lower bounded by $L_k = \tau_k^{-1} + \alpha_k L_{\mathbf{A}}$ and $\sigma_k = \tau_k^{-1} - \mathbf{u}_k^{\mathbf{H}} \mathbf{u}_k / \rho_k^{\mathbf{B}}$, respectively, where Lemma 4 shows that $\sigma_k > 0$. Algorithm 3 describes the accelerated projection method with fixed momentum for solving (11).

III. CONVERGENCE ANALYSIS

This section provides the convergence analysis of Algorithm 1 for solving (6). Before presenting the convergence

Algorithm 3 Accelerated Projection Method with Fixed Momentum for Computing the WPM

Initialization: $\bar{\mathbf{x}}_1 = \mathbf{x}_k$, $\mathbf{z}_1 = \mathbf{x}_k$, tolerance $\epsilon = 10^{-6}$, stepsize $\frac{1}{L_k}$ and $\kappa_k = \frac{L_k}{\sigma_k}$ where L_k (respectively, σ_k) denotes the upper (respectively, lower) bound of the largest (respectively, smallest) eigenvalue of $\mathbf{B}_k + \alpha_k \mathbf{A}^{\mathbf{H}} \mathbf{A}$

Iteration:

- 1: **for** $i = 1, 2, \dots$ **do**
 - 2: $\bar{\mathbf{x}}_{i+1} \leftarrow \text{prox}_{L_k}^{\mathbf{I}}(\mathbf{z}_i - \frac{1}{L_k} \nabla G_k(\mathbf{z}_i))$
 - 3: **if** $\|\bar{\mathbf{x}}_{i+1} - \bar{\mathbf{x}}_i\| \leq \epsilon$ **then**
 - 4: break
 - 5: **else**
 - 6: $\mathbf{z}_{i+1} \leftarrow \bar{\mathbf{x}}_{i+1} + \frac{\sqrt{\kappa_k} - 1}{\sqrt{\kappa_k} + 1} (\bar{\mathbf{x}}_{i+1} - \bar{\mathbf{x}}_i)$
 - 7: **end if**
 - 8: **end for**
-

results, we first introduce two assumptions and then provide four lemmas that simplify presenting the convergence proof.

Assumption 1 (L -Smooth regularizer). *We assume the gradient of $f(\mathbf{x})$ is L -Lipschitz continuous, meaning that for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^N$, there exists a positive constant L such that the following inequality holds:*

$$\|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2)\| \leq L \|\mathbf{x}_1 - \mathbf{x}_2\|. \quad (14)$$

Assumption 2 (Constrained proximal Polyak-Łojasiewicz inequality [73, 76]). *Define*

$$\begin{aligned} S_h(\bar{\mathbf{x}}, \mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha) &\triangleq \Re\{\langle \mathbf{g}, \bar{\mathbf{x}} - \mathbf{x} \rangle\} + \frac{1}{2\alpha} \|\bar{\mathbf{x}} - \mathbf{x}\|_{\mathbf{W}}^2 \\ &\quad + h(\bar{\mathbf{x}}) - h(\mathbf{x}), \end{aligned} \quad (15)$$

and

$$\mathcal{D}_h^{\mathcal{C}}(\mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha) \triangleq -\frac{2}{\alpha} \min_{\bar{\mathbf{x}} \in \mathcal{C}} S_h(\bar{\mathbf{x}}, \mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha), \quad (16)$$

where $\mathbf{W} \in \mathbb{C}^{N \times N}$, $\mathbf{W} \succ 0$ is a Hermitian positive definite matrix, h was defined in (6) and α is a positive constant. If there exists a positive constant ν such that the following inequality holds:

$$\mathcal{D}_h^{\mathcal{C}}(\mathbf{x}, \nabla f(\mathbf{x}), \mathbf{I}_N, \alpha) \geq 2\nu(F(\mathbf{x}) - F^*), \quad \forall \mathbf{x} \in \mathcal{C}, \quad (17)$$

then we say $F(\mathbf{x})$ satisfies the constrained proximal Polyak-Łojasiewicz inequality. F^* denotes the optimal value of (6).

Assumption 1 is a moderate assumption and is commonly used in convergence analysis for differentiable functions, see [57, 58]. We analysed the validity of Assumption 1 for the used network architecture (see Fig. 3), along with empirical validation in the supplementary material. Assumption 2 covers certain nonconvex properties of $F(\mathbf{x})$ and is commonly used in nonconvex analysis [76]. Next we introduce Lemmas 1 to 4 which are useful for the following convergence proofs.

Lemma 1 (Majorizer of f). *Let $f : \mathbb{C}^N \rightarrow (-\infty, \infty]$ be an L -smooth function ($L > 0$). Then for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^N$, we have*

$$f(\mathbf{x}_2) \leq f(\mathbf{x}_1) + \Re\{\langle \nabla f(\mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1 \rangle\} + \frac{L}{2} \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2. \quad (18)$$

Lemma 2. For any fixed $\mathbf{W} \in \mathbb{C}^{N \times N}$, $\mathbf{W} \succ 0$ and $\alpha > 0$, we have the following inequality

$$\mathcal{D}_h^C(\mathbf{x}, \nabla f(\mathbf{x}), \mathbf{W}, \alpha) \geq \|\mathcal{G}_{\frac{1}{\alpha}, \mathbf{W}}^{f, h}(\mathbf{x})\|_{\mathbf{W}}^2, \quad \forall \mathbf{x} \in \mathcal{C},$$

where $\mathcal{G}_{\frac{1}{\alpha}, \mathbf{W}}^{f, h}(\mathbf{x}) \triangleq \frac{1}{\alpha}(\mathbf{x} - \bar{\mathbf{x}}^+)$ with

$$\bar{\mathbf{x}}^+ \triangleq \arg \min_{\bar{\mathbf{x}} \in \mathcal{C}} \mathcal{S}_h(\bar{\mathbf{x}}, \mathbf{x}, \nabla f(\mathbf{x}), \mathbf{W}, \alpha). \quad (19)$$

Lemma 3. For any fixed $\mathbf{W} \in \mathbb{C}^{N \times N}$, $\mathbf{W} \succ 0$, a differentiable function f and a convex function h , $\mathcal{D}_h^C(\mathbf{x}, \nabla f(\mathbf{x}), \mathbf{W}, \alpha)$ satisfies the following inequality for $\forall \alpha_1 \geq \alpha_2 > 0$:

$$\mathcal{D}_h^C(\mathbf{x}, \nabla f(\mathbf{x}), \mathbf{W}, \alpha_2) \geq \mathcal{D}_h^C(\mathbf{x}, \nabla f(\mathbf{x}), \mathbf{W}, \alpha_1). \quad (20)$$

Lemma 4 (Bounded Hessian). The approximate Hessian matrices generated by Algorithm 2 satisfy the following inequality

$$\underline{\eta} \mathbf{I} \preceq \mathbf{B}_k \preceq \bar{\eta} \mathbf{I},$$

where $0 < \underline{\eta} < \bar{\eta} < \infty$.

The proof of Lemma 1 is similar to the one in the real-valued case [74]. The only difference is the use of $\Re\{\langle \nabla f(\mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1 \rangle\}$ instead of $\langle \nabla f(\mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1 \rangle$ to account for the complex domain. Therefore, we omit the proof here for brevity. Appendices A to C give the proofs of Lemmas 2 to 4. With these in place, we are able to derive our main convergence results in Theorem 1.

Theorem 1 (Convergence results). Applying Algorithm 1 to solve (6), we can establish the following convergence results:

- Let $\alpha_k \leq \frac{\eta}{L}$, and define $\Delta_K \triangleq \min_{k \leq K} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2$. Under Assumption 1, by running Algorithm 1 K iterations to solve (6), we have

$$\Delta_K \leq \frac{2(F(\mathbf{x}_1) - F^*)}{LK},$$

where F^* denotes the minimum of (6) and $\mathbf{x}_1$ is the initial iterate.

- Let $\alpha_k \leq \min\left\{\frac{\eta}{\nu}, \frac{\eta}{L}\right\}$. Under Assumptions 1 and 2, we have the following convergence rate bound for the cost value

$$F(\mathbf{x}_{k+1}) - F^* \leq \left(1 - \frac{\nu \alpha_{\min}}{\bar{\eta}}\right)^k (F(\mathbf{x}_1) - F^*),$$

where $\alpha_{\min} = \min_k \{\alpha_k\}$.

- Let $\alpha_k \leq \frac{\eta}{L}$. Choose $k \in \{1, \dots, K-1\}$ uniformly at random. Then under Assumptions 1 and 2, we have the following convergence rate bound in expectation

$$\mathbb{E}[F(\mathbf{x}_k) - F^*] \leq \frac{\bar{\eta}(F(\mathbf{x}_1) - F^*)}{\nu \alpha_{\min} K}.$$

Appendix D presents the proof. The first part of Theorem 1 shows that $\Delta_K \rightarrow 0$ as $K \rightarrow \infty$. Combining with the summation in (26) implies convergence to a fixed point. The second part establishes that the cost function sequence converges linearly to the minimal cost value. The third part demonstrates that if one selects a random iterate, then the cost value converges sublinearly in expectation. In this paper, we

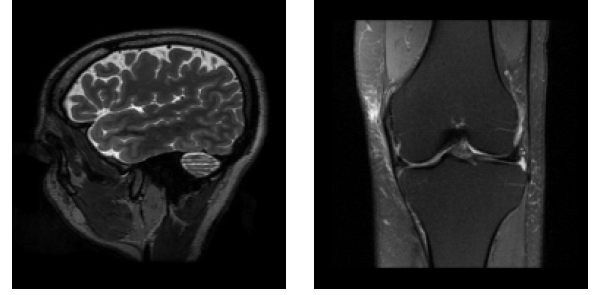
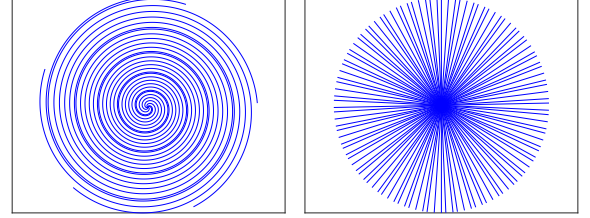
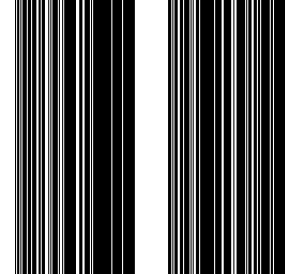


Fig. 1. Magnitude of the complex-valued ground-truth images.



(a) Spiral (b) Radial



(c) Cartesian Sampling Mask

Fig. 2. The spiral (a) and radial (b) sampling trajectories and the Cartesian sampling mask (c) used in the experiments.

simply choose the last iterate as the output. However, Section IV-A discusses the third term experimentally to validate our analysis. Note that Assumption 2 is a special case of the Kurdyka-Łojasiewicz (KL) inequality with exponent $1/2$ [77]. While our analysis establishes convergence rate under the PL inequality, it could potentially be extended to the more general KL inequality, see [78]. Moreover, if the Dennis-Moré condition [79] holds, one can expect a superlinear convergence rate. However, a full study of this direction is beyond the scope of the present paper and is left for future work.

IV. NUMERICAL EXPERIMENTS

This section investigates the performance of the proposed method for CS MRI reconstruction using spiral, radial, and Cartesian k-space sampling patterns. We first describe the experimental and algorithmic settings. Then, we present the reconstruction results and experimentally demonstrate the convergence of the proposed method, validating our theoretical analysis.

Experimental Settings: We evaluated the performance of the proposed method using brain and knee MRI datasets. For the

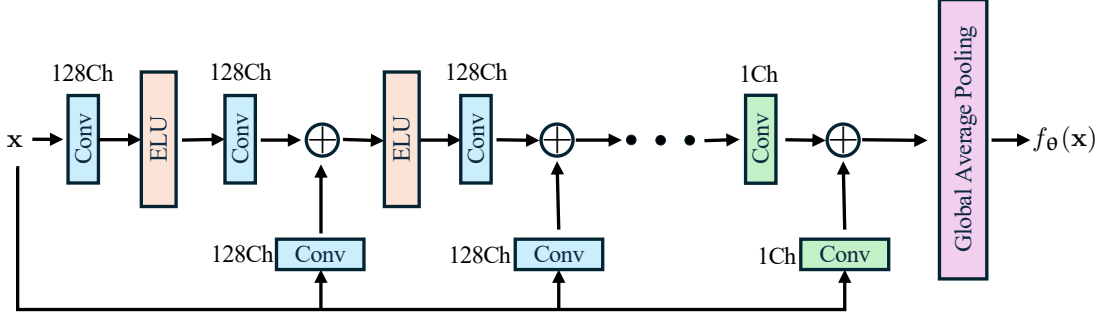


Fig. 3. The neural network architecture used for the energy function $f_{\theta}(\mathbf{x})$, based on [57]. The convolutional kernel size is 3×3 with stride 1.

brain images, we used the dataset from [17], which consists of 360 images for training and 164 images for testing. For knee images, we employed the NYU fastMRI [80] multi-coil knee dataset. The ESPIRiT algorithm [81] was first applied to reconstruct complex-valued images from multi-coil k-space data. All brain and knee images were cropped and resized to a resolution of 256×256 and normalized such that their maximum magnitude is one. We employed the network architecture proposed in [57] to construct $f(\mathbf{x})$. Unlike [57], we included bias terms in the network, as this provided improved performance in our experiments. For completeness, Fig. 3 presents the used network architecture. The number of layers is set to six, which is identical to that in [57]. Noisy images were generated by adding independent and identically distributed (i.i.d.) Gaussian noise with a variance of $1/255$. The network was trained to enforce $\mathbf{x} - \nabla f(\mathbf{x})$ to be a denoiser and $\nabla f(\mathbf{x})$ was computed with PyTorch's autograd function, i.e., `torch.autograd.grad`. The training process used the mean squared error as the loss function, with a batch size of 64. Optimization was performed using the ADAM algorithm [82] with a learning rate 10^{-3} over a total of 18,000 iterations. Additionally, we halved the learning rate at each 4,000 iterations. Although we trained distinct denoisers for brain and knee images, the same denoiser was employed across different sampling acquisitions. Note that training denoisers on larger datasets may lead to improved denoisers and then better reconstruction. However, the main focus of this paper is to investigate numerical efficiency of gradient-driven denoisers based CS MRI reconstruction with convergence guarantees. Thus, we leave the exploration of this direction to future work.

To evaluate reconstruction performance, we selected 6 brain and 6 knee images from the test datasets as the ground-truth. Fig. 1 presents the magnitudes of two of these twelve complex-valued ground truth images. Due to space limitations, the remaining ten brain and knee ground-truth images are presented in the supplementary material. For the spiral trajectory, we used 6 interleaves, 1688 readout points, and 32 coils, whereas the radial trajectory involved 55 spokes with golden-angle rotation, 1024 readout points, and 32 coils. Fig. 2 illustrates the sampling trajectories and mask used in our experiments. To generate the k-space data, we first applied the forward model with the corresponding trajectories and sensitivity maps to the ground-truth images. We then added complex i.i.d. Gaussian

noise with zero mean and a variance of 10^{-4} , resulting in noisy measurements with an input SNR of approximately 21dB. In the reconstruction step, we used coil compression [83] to reduce the 32 coils to 12 virtual coils in order to reduce computational complexity. Sections IV-A to IV-C specifically analyze one brain and one knee test image acquired using the spiral, radial, and Cartesian sampling trajectories. Additional results on other test images are provided in the supplementary material. All experiments were implemented using PyTorch and executed on an NVIDIA A100 GPU.

Algorithmic Settings: We primarily compared our method with the projected gradient descent method (dubbed GD) [57] and the proximal gradient method (dubbed PG) [58], as both methods also come with convergence guarantees, similar to ours. In addition, we compared with the accelerated proximal gradient method (dubbed APG) [84] that also provides convergence guarantees for solving nonconvex problems such as (6). Since we normalized the maximum magnitude of all images to one, we set $\mathcal{C} = \{\mathbf{x} \mid \|\mathbf{x}\|_{\infty} \leq 1\}$ and all methods used this constraint. Although such a constraint is not needed in practical CS MRI reconstruction, we considered it here primarily to test the applicability of CQNPM to constrained problems. Following [85], which was also adopted in PnP-LBFGS, we attempted to train a network to meet the requirements of PnP-LBFGS for CS MRI reconstruction. However, we observed that PnP-LBFGS diverged. Therefore, we compared our method against PnP-LBFGS on the image deblurring problem using the same experimental setup in their open-source code. The results, provided in the supplementary material, show that our method converged faster than PnP-LBFGS in terms of PSNR with respect to both iterations and wall time.

For plots involving F^* , we ran APG for 500 iterations and set $F^* = F(\mathbf{x}_{500})$. The stepsize α_k in Algorithm 1 is set to be one which we found it works well for our experimental settings. Alternatively, one can choose α_k with a backtracking line search strategy to satisfy (22). Algorithm 2 used $\delta = 10^{-8}$, $\theta_1 = 2 \times 10^{-6}$, $\theta_2 = 200$. The parameter δ is commonly used in quasi-Newton methods to guard numerical stability when $\mathbf{v}_k$ and $\mathbf{s}_k$ are closely related. Parameters θ_1 and θ_2 control the Hermitian positive definiteness of the estimated Hessian matrices but still allow greater flexibility of estimation. In view of the proof in Lemma 4, we have $\frac{1}{2\theta_2} \mathbf{I}_N \preceq \mathbf{H}_k \preceq \left(\frac{1+\delta}{\delta\theta_1}\right) \mathbf{I}_N$. So θ_1 and θ_2 should be positive,

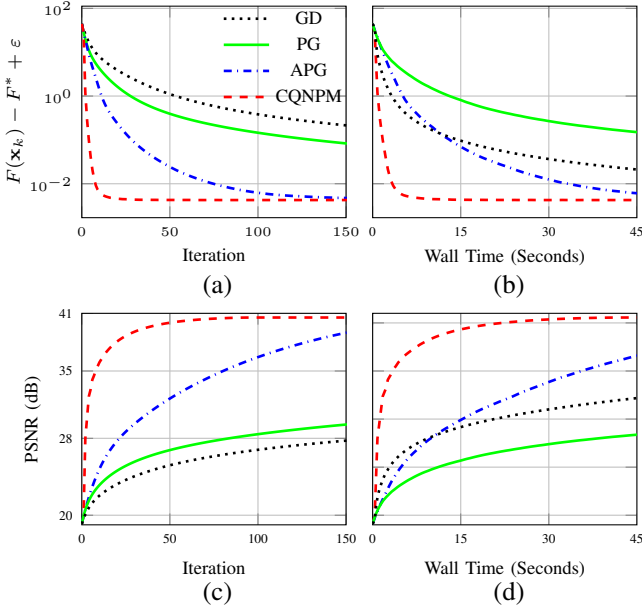


Fig. 4. Comparison of different methods with spiral acquisition on the brain image for $\varepsilon = 4 \times 10^{-3}$. (a), (b): cost values versus iteration and wall time; (c), (d): PSNR values versus iteration and wall time.

with θ_1 chosen small and θ_2 chosen large. In Algorithm 3, we always used the previous iterate as the new initial value. Therefore, we found that setting the maximum number of iterations to 15 and $\varepsilon = 10^{-6}$ worked well in practice.

A. Spiral Acquisition Reconstruction

Fig. 4 presents the cost and PSNR values of each method with respect to the number of iterations and wall time for the reconstruction with spiral acquisition on the brain image. Figs 4 (a) and (c) show that our method converges faster than the others, reaching a lower cost value and higher PSNR at the same number of iterations. Moreover, we observed that APG is faster than PG, and PG is faster than GD in terms of the number of iterations, which aligns with our expectations. From Figs 4 (b) and (d), we saw that our method is also the fastest algorithm in terms of wall time. Moreover, we observed that GD becomes faster than PG in terms of wall time because PG needs to solve a proximal mapping requiring to execute $\mathbf{Ax}$ multiple times. However, APG is still faster than GD even it also requires to solve a proximal mapping.

Fig. 5 shows the reconstructed images at the 50th and 100th iterations. Our method recovered significantly clearer images than other methods at the same number of iterations. Moreover, the corresponding error maps clearly show that our method yielded small reconstruction errors. The supplementary material presents the reconstruction results of the knee image using the same spiral acquisition, as well as results for other additional test images, where similar trends were observed across all methods.

B. Radial Acquisition Reconstruction

Fig. 6 summarizes the cost and PSNR values of each approach with respect to the number of iterations and wall

time for the reconstruction with radial acquisition on the knee image. It is apparent that the performance trends here are consistent with those observed in the reconstruction with spiral acquisition on the brain image. Fig. 7 presents the reconstructed images and error maps of each method at the 50 and 100th iterations. It is evident that our method outperformed other methods. The supplementary material summarizes the results of other additional test images, where similar trends were observed.

C. Cartesian Acquisition Reconstruction

Fig. 8 shows the cost and PSNR values of each method with respect to the number of iterations and wall time for the reconstruction with Cartesian acquisition on the brain image. It is evident that similar trends were identified. Moreover, we observed that the difference between GD and PG in terms of wall time was smaller than in the cases of spiral and radial acquisitions. This is because computing $\mathbf{Ax}$ is less expensive in this case than in the non-Cartesian acquisitions, resulting in a more efficient evaluation of the proximal mapping. Fig. 9 shows the reconstructed images and error maps at the 50 and 100th iterations. There is no doubt that our method yielded a clearer image than the others. The supplementary material shows additional experimental results of five other brain images, where similar trends were obtained.

D. Convergence Validation

We studied the convergence properties of our method experimentally. Let $E(\mathbf{x}_k) = \|\mathbf{x}_k - \mathbf{x}_{k+1}\|_2^2$. Fig. 10 shows the cost values and the normalized difference $E(\mathbf{x}_k)/E(\mathbf{x}_1)$ of our method with spiral acquisition on all brain test images. As expected, the cost values converged to a constant for all test images, and $E(\mathbf{x}_k)/E(\mathbf{x}_1) \rightarrow 0$, consistent with our theoretical analysis. Fig. 11 presents the expected cost values versus iteration, estimated using the Monte Carlo method with 1000 samples. We observed that the expected cost values decreased similarly to Fig. 10 (a), though slightly more slowly, aligning well with our theoretical results.

E. Comparison with the Preconditioned PnP Approach

We studied the difference between using gradient-driven denoisers and classical CNN-based denoisers within the PnP framework. Specifically, we adopted the preconditioned PnP method (P²nP) [37], as it generally outperforms the classical PnP-ISTA method. Fig. 12 reports the PSNR values versus iterations for CQNPM and P²nP under spiral (respectively, radial) acquisition for the brain (respectively, knee) image. We observed that CQNPM converged faster than P²nP and consistently achieved higher PSNR values, illustrating that the use of gradient-driven denoisers does not sacrifice reconstruction performance.

V. CONCLUSION AND FUTURE WORK

Compared to the PnP/RED framework with convolutional neural network based denoisers, the gradient-driven denoisers offer a much stronger theoretical foundation, as its required

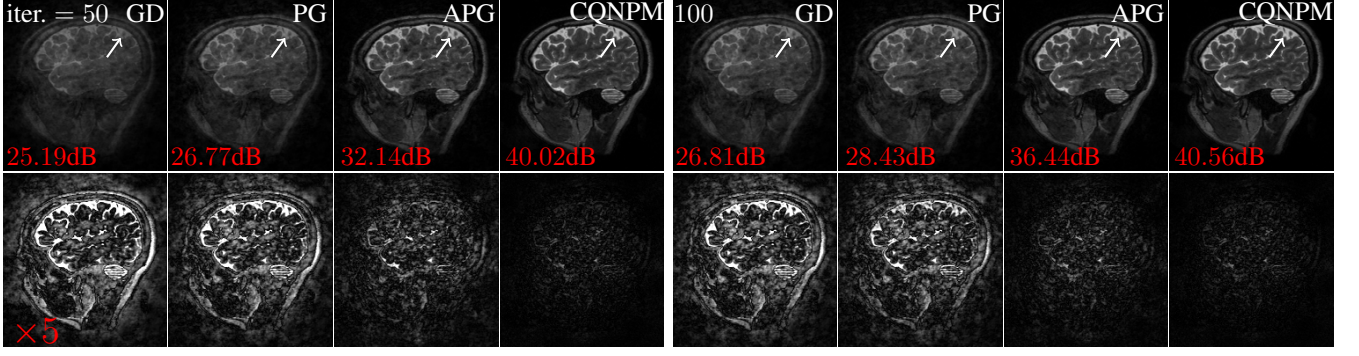


Fig. 5. First row: the reconstructed brain images of each method at 50 and 100th iterations with spiral acquisition. The PSNR values are labeled at the left bottom corner of each image. Second row: the associated error maps ($\times 5$) of the reconstructed images.

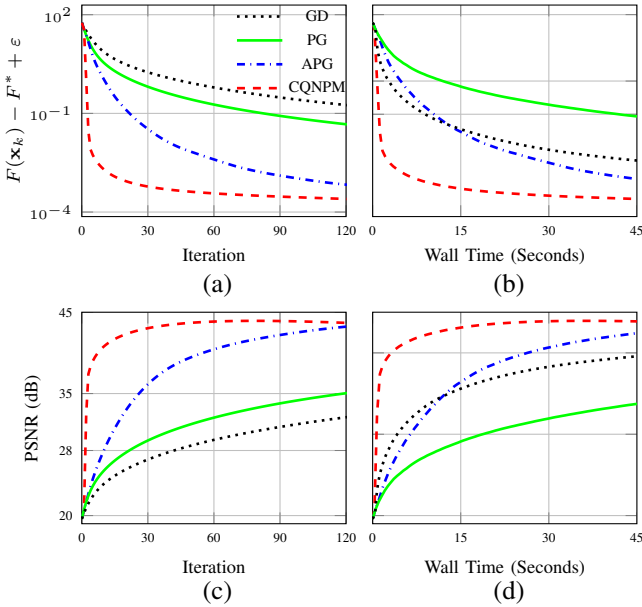


Fig. 6. Comparison of different methods with radial acquisition on the knee image for $\varepsilon = 8 \times 10^{-5}$. (a), (b): cost values versus iteration and wall time; (c), (d): PSNR values versus iteration and wall time.

assumption (i.e., Lipschitz continuity of $f(\mathbf{x})$) is easier to satisfy in practice—an important advantage for medical imaging applications. We applied the gradient-driven denoisers to CS MRI reconstruction and thoroughly evaluated its performance on spiral, radial, and Cartesian acquisitions. In addition, we proposed a *complex* quasi-Newton proximal method (CQNPM) to efficiently solve the associated minimization problem with convergence guarantee under nonconvex settings. We extensively compared our method with existing algorithms, and the experimental results demonstrate both the efficiency of our approach and the accuracy of the underlying theoretical analysis. Although we consider $h(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2$ in this paper, the algorithm presented is applicable to any h that is convex and smooth (Lipschitz gradient). Furthermore, our theoretical results require only the convexity of h , so the approach also generalizes to non-smooth and convex h by modifying the weighted proximal mapping at step 3 in

Algorithm 1 appropriately. Code to reproduce the results in the paper is available at <https://github.com/hongtao-argmin/CQNPM-GDD-CS-MRI-Reco>.

While the proposed CQNPM accelerates convergence in gradient-driven denoisers based CS MRI reconstruction, several limitations remain. In the following, we outline these limitations and discuss possible directions for improvement:

- **Accurate Hermitian positive definite Hessian Approximation:** In this work, we employed Algorithm 2 to estimate a Hermitian positive definite Hessian matrix, but we always reinitialized the Hessian at each step. In the real-valued setting, BFGS updating typically provides more accurate Hessian approximations compared to the scheme in Algorithm 2. Therefore, extending BFGS to our problem—while ensuring that the Hessian matrix remains Hermitian positive definite when f is nonconvex—could potentially yield a faster algorithm.
- **Assumptions for convergence:** The convergence rate of our method depends on the proximal PL inequality. However, verifying this condition for complex, high-capacity neural network-based denoisers remains challenging in practice. Extending the analysis to the more general KL inequality, or alternatively designing convex networks [86] that still achieve comparable performance, would be promising directions for future research.
- **Extension to 3D and dynamic MRI reconstruction:** This work focuses on 2D multi-coil CS MRI reconstruction. Extending CQNPM to 3D or dynamic MRI reconstruction would be an interesting future direction. However, training effective gradient-driven denoisers for such high-dimensional problems remains a significant challenge that must be addressed.

APPENDIX A PROOF OF LEMMA 2

Since $\mathbf{W} \succ 0$ and $h(\mathbf{x})$ is convex, $\mathcal{S}_h(\bar{\mathbf{x}}, \mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha)$ is strongly convex. Through the optimality condition of (19), we have the following inequality

$$\Re \left\{ \left\langle \mathbf{g} + \frac{1}{\alpha} \mathbf{W}(\bar{\mathbf{x}}^+ - \mathbf{x}) + \nabla h(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}}^+ \right\rangle \right\} \geq 0, \quad \forall \mathbf{x} \in \mathcal{C}.$$

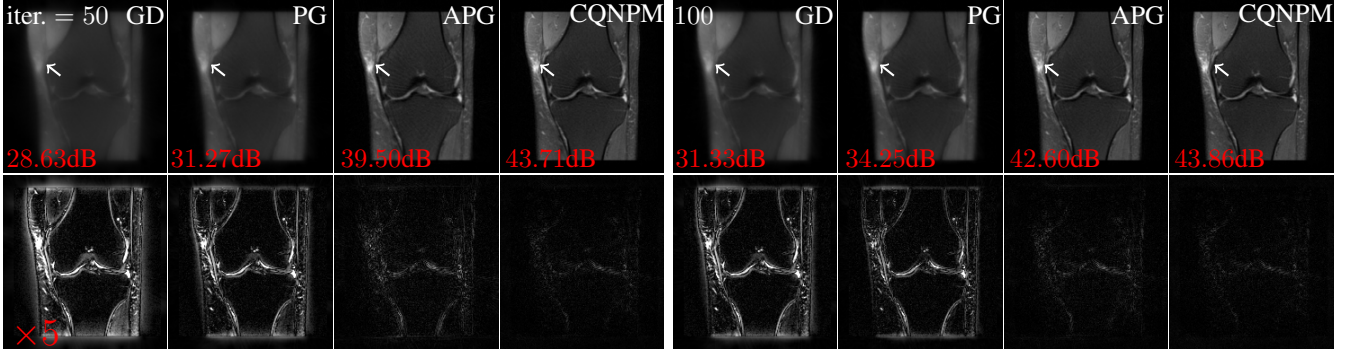


Fig. 7. First row: the reconstructed knee images of each method at 50 and 100th iterations with radial acquisition. The PSNR values are labeled at the left bottom corner of each image. Second row: the associated error maps ($\times 5$) of the reconstructed images.

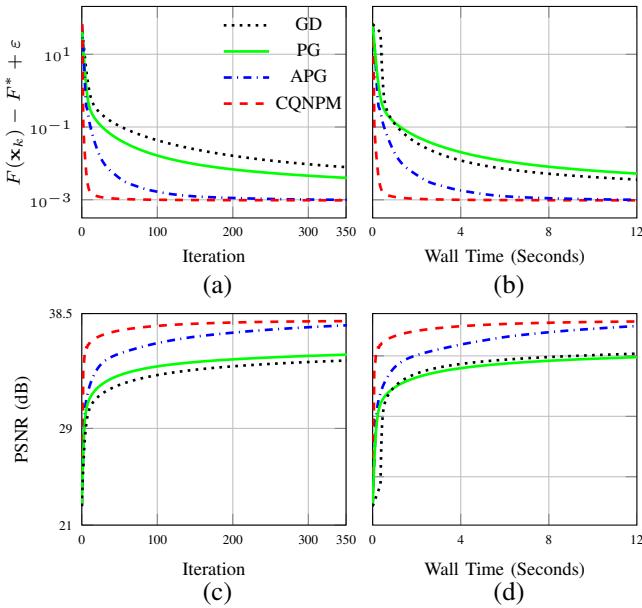


Fig. 8. Comparison of different methods with Cartesian acquisition on the brain image for $\varepsilon = 9 \times 10^{-4}$. (a), (b): cost values versus iteration and wall time; (c), (d): PSNR values versus iteration and wall time.

By using the convexity of h (i.e., $h(\mathbf{x}) - h(\bar{\mathbf{x}}) \geq \Re\{\langle \nabla h(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle\}$), we reach

$$\Re\{\langle \mathbf{g}, \mathbf{x} - \bar{\mathbf{x}}^+ \rangle\} \geq \frac{1}{\alpha} \|\bar{\mathbf{x}}^+ - \mathbf{x}\|_{\mathbf{W}}^2 + h(\mathbf{x}^+) - h(\mathbf{x}). \quad (21)$$

From (16) and (19), we have

$$\begin{aligned} \mathcal{D}_h^c(\mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha) &= -\frac{2}{\alpha} \mathcal{S}_h(\bar{\mathbf{x}}^+, \mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha) \\ &= \frac{2}{\alpha} \Re\{\langle \mathbf{g}, \mathbf{x} - \bar{\mathbf{x}}^+ \rangle\} - \frac{1}{\alpha^2} \|\bar{\mathbf{x}}^+ - \mathbf{x}\|_{\mathbf{W}}^2 \\ &\quad - \frac{2}{\alpha} (h(\bar{\mathbf{x}}^+) - h(\mathbf{x})) \\ &\geq \frac{1}{\alpha^2} \|\bar{\mathbf{x}}^+ - \mathbf{x}\|_{\mathbf{W}}^2 \\ &= \|\mathcal{G}_{\frac{1}{\alpha}}^{f,h}(\mathbf{x})\|_{\mathbf{W}}^2, \end{aligned}$$

where the first inequality and last equality come from (21) and the definition of $\mathcal{G}_{\frac{1}{\alpha}}^{f,h}(\mathbf{x})$.

APPENDIX B PROOF OF LEMMA 3

Let $\alpha_1 \geq \alpha_2 > 0$ and $\bar{\mathbf{x}}_2 = \frac{\alpha_2}{\alpha_1} \bar{\mathbf{x}}_1 + \frac{\alpha_1 - \alpha_2}{\alpha_1} \mathbf{x}$ with $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2 \in \mathcal{C}$. Since $h(\mathbf{x})$ is convex, we get

$$h(\bar{\mathbf{x}}_2) \leq \frac{\alpha_2}{\alpha_1} h(\bar{\mathbf{x}}_1) + \frac{\alpha_1 - \alpha_2}{\alpha_1} h(\mathbf{x})$$

and then

$$\alpha_1 (h(\bar{\mathbf{x}}_2) - h(\mathbf{x})) \leq \alpha_2 (h(\bar{\mathbf{x}}_1) - h(\mathbf{x})).$$

By using $\alpha_1 (\bar{\mathbf{x}}_2 - \mathbf{x}) = \alpha_2 (\bar{\mathbf{x}}_1 - \mathbf{x})$ and the above inequality, we have

$$-\frac{2}{\alpha_1} \mathcal{S}_h(\bar{\mathbf{x}}_1, \mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha_1) \leq -\frac{2}{\alpha_2} \mathcal{S}_h(\bar{\mathbf{x}}_2, \mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha_2).$$

Minimizing both sides of the above inequalities and using the definition of $\mathcal{D}_h^c(\mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha)$, we get the desired result

$$\mathcal{D}_h^c(\mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha_1) \leq \mathcal{D}_h^c(\mathbf{x}, \mathbf{g}, \mathbf{W}, \alpha_2).$$

APPENDIX C PROOF OF LEMMA 4

To prove the bound on $\mathbf{B}_k$, we first establish the bound on $\mathbf{H}_k$ because $\mathbf{B}_k = \mathbf{H}_k^{-1}$. Define $a_k = \langle \mathbf{s}_k, \mathbf{s}_k \rangle$, $b_k = \Re\{\langle \mathbf{s}_k, \mathbf{v}_k \rangle\}$, $c_k = \langle \mathbf{v}_k, \mathbf{v}_k \rangle$. Then we have

$$\frac{b_k}{a_k} \tau_k = 1 - \sqrt{1 - \frac{b_k^2}{a_k c_k}} \leq \frac{b_k^2}{a_k c_k},$$

resulting in $\tau_k \leq \frac{b_k}{c_k}$. The last inequality follows from the Cauchy-Schwarz inequality, which leads to $b_k^2 \leq a_k c_k$. The lower bound can be derived through

$$\begin{aligned} \tau_k &= \frac{a_k}{b_k} - \sqrt{\left(\frac{a_k}{b_k}\right)^2 - \frac{a_k}{c_k}} = \frac{\frac{a_k}{c_k}}{\frac{a_k}{b_k} + \sqrt{\left(\frac{a_k}{b_k}\right)^2 - \frac{a_k}{c_k}}} \\ &> \frac{b_k}{2c_k}. \end{aligned}$$

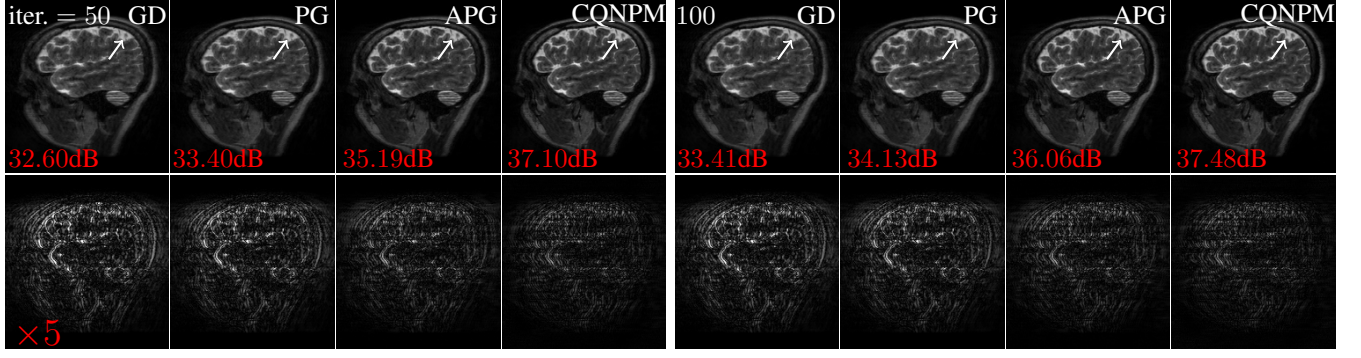


Fig. 9. First row: the reconstructed brain images of each method at 50 and 100th iterations with Cartesian acquisition. The PSNR values are labeled at the left bottom corner of each image. Second row: the associated error maps ($\times 5$) of the reconstructed images.

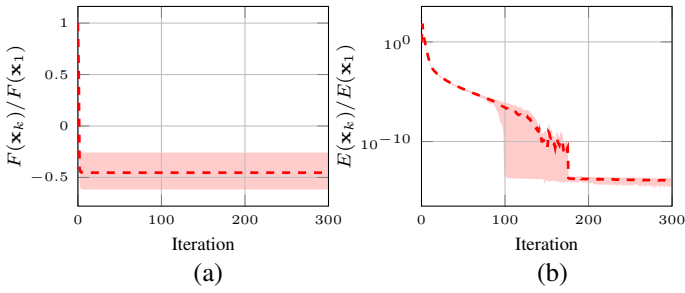


Fig. 10. Averaged cost values (a) and $E(\mathbf{x}_k)/E(\mathbf{x}_1)$ (b) versus iteration for the proposed method. The shaded region of each curve represents the range of the cost values and $E(\mathbf{x}_k)$ across six brain test images with spiral acquisition.

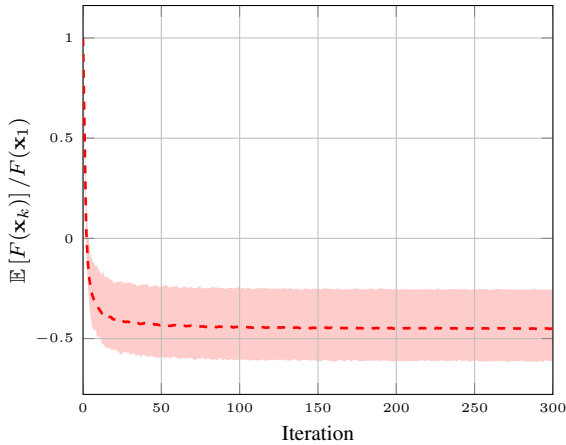


Fig. 11. Averaged expected cost values versus iteration for the proposed method. The shaded region of each curve represents the range of expected cost values across six brain test images with spiral acquisition. The cost values were estimated using the Monte Carlo method with 1000 samples.

In summary, we have $\frac{b_k}{2c_k} \leq \tau_k \leq \frac{b_k}{c_k}$. Next, we derive an upper bound for $\mathbf{u}_k^H \mathbf{u}_k$. Using the definition of $\mathbf{u}_k$, we have

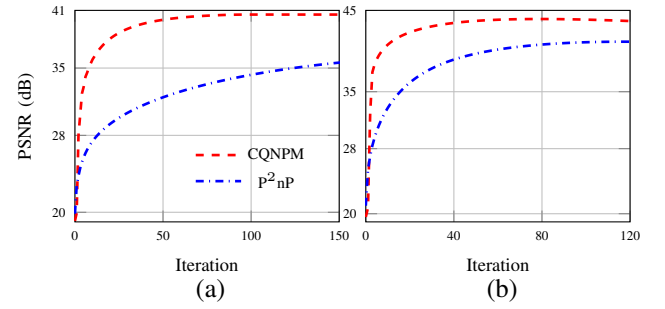


Fig. 12. Comparison of the preconditioned PnP (P^2PnP) method. (a) PSNR values versus iteration for spiral acquisition with the brain image; (b) PSNR values versus iteration for radial acquisition with the knee image.

the following inequalities

$$\begin{aligned}
 \mathbf{u}_k^H \mathbf{u}_k &= \frac{\langle \mathbf{s}_k - \tau_k \mathbf{v}_k, \mathbf{s}_k - \tau_k \mathbf{v}_k \rangle}{\rho_k} \\
 &\leq \frac{\|\mathbf{s}_k - \tau_k \mathbf{v}_k\|_2}{\delta \|\mathbf{v}_k\|_2} \\
 &= \frac{1}{\delta} \sqrt{\frac{a_k}{c_k} - \frac{2\tau_k b_k}{c_k} + \tau_k^2} \\
 &\leq \frac{1}{\delta} \sqrt{\frac{a_k}{c_k}} = \frac{1}{\delta} \sqrt{\frac{a_k}{b_k} \frac{b_k}{c_k}} \leq \frac{a_k}{\delta b_k}.
 \end{aligned}$$

The first inequality comes from step 5 in Algorithm 2. The second inequality derived from the fact that $\frac{b_k}{2c_k} \leq \tau_k \leq \frac{b_k}{c_k}$ resulting in $-\frac{2\tau_k b_k}{c_k} + \tau_k^2 < 0$. The last inequality is the result of $b_k^2 \leq a_k c_k$.

With these, we have $\frac{b_k}{2c_k} \mathbf{I}_N \preceq \mathbf{H}_k \preceq (\frac{b_k}{c_k} + \frac{a_k}{\delta b_k}) \mathbf{I}_N$. From (13), we know $\frac{b_k}{a_k} \geq \theta_1$ and $\frac{c_k}{b_k} \leq \theta_2$ for all k . Therefore, $\mathbf{H}_k$ is always bounded by $\frac{1}{2\theta_2} \mathbf{I}_N \preceq \mathbf{H}_k \preceq (\frac{1+\delta}{\delta\theta_1}) \mathbf{I}_N$. Since $\mathbf{H}_k$ is both lower and upper bounded, we know that there exist constants $\bar{\eta} > \underline{\eta} > 0$ such that $\underline{\eta} \mathbf{I} \preceq \mathbf{B}_k \preceq \bar{\eta} \mathbf{I}$.

APPENDIX D
PROOF OF THEOREM 1

Using Lemma 1, we have

$$\begin{aligned}
f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \Re\{\langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle\} \\
&\quad + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \\
&\leq f(\mathbf{x}_k) + h(\mathbf{x}_k) - h(\mathbf{x}_{k+1}) \\
&\quad + \min_{\mathbf{v} \in \mathcal{C}} \mathcal{S}_h(\mathbf{v}, \mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) \\
&\quad + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 - \frac{1}{2\alpha_k} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{B}_k}^2 \\
&\leq F(\mathbf{x}_k) + \left(\min_{\mathbf{v} \in \mathcal{C}} \mathcal{S}_h(\mathbf{v}, \mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) \right) \\
&\quad + \left(\frac{L}{2} - \frac{\eta}{2\alpha_k} \right) \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 - h(\mathbf{x}_{k+1}).
\end{aligned}$$

The second inequality comes from the definition of $\mathcal{S}_h$ and the update rule for $\mathbf{x}_{k+1}$. The third inequality is the result of Lemma 4. Letting $\alpha_k \leq \frac{\eta}{L}$ and moving $h(\mathbf{x}_{k+1})$ to the left side, we get

$$\begin{aligned}
F(\mathbf{x}_{k+1}) &\leq F(\mathbf{x}_k) + \left(\min_{\mathbf{v} \in \mathcal{C}} \mathcal{S}_h(\mathbf{v}, \mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) \right) \\
&\quad + \left(\frac{L}{2} - \frac{\eta}{2\alpha_k} \right) \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \\
&\leq F(\mathbf{x}_k) - \frac{\alpha_k}{2} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k). \quad (22)
\end{aligned}$$

Rearranging (22), we have

$$\frac{\alpha_k}{2} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) \leq F(\mathbf{x}_k) - F(\mathbf{x}_{k+1}). \quad (23)$$

Invoking Lemmas 2 and 4, we get

$$\mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) \geq \frac{\eta}{\alpha_k} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2. \quad (24)$$

Substituting (24) into (23), we have

$$\frac{\eta}{2\alpha_k} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \leq F(\mathbf{x}_k) - F(\mathbf{x}_{k+1}). \quad (25)$$

Summing up the above inequality from $k = 1$ to K , we reach

$$\sum_{k=1}^K \frac{\eta}{2\alpha_k} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \leq F(\mathbf{x}_1) - F(\mathbf{x}_K) \leq F(\mathbf{x}_1) - F^*. \quad (26)$$

Denote by $\alpha_{\max} = \max_k \{\alpha_k\}$, $\alpha_{\min} = \min_k \{\alpha_k\}$, and $\Delta_K = \min_{k \leq K} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2$. Since $\alpha_k \leq \frac{\eta}{L}$, we have $\alpha_{\max} = \frac{\eta}{L}$. Invoking the definition of Δ_K , the value of $\alpha_{\max}$, and using (26), we obtain

$$\Delta_K \leq \frac{2(F(\mathbf{x}_1) - F^*)}{LK}.$$

Clearly, Δ_K approaches zero as $K \rightarrow \infty$.

Now, we show the convergence of cost values with additional Assumption 2. Invoking Lemmas 3 and 4, we have the following inequalities

$$\begin{aligned}
\mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) &\geq \frac{1}{\eta} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{I}, \frac{\alpha_k}{\eta}) \\
&\geq \frac{1}{\eta} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{I}, \frac{\alpha_{\max}}{\eta}). \quad (27)
\end{aligned}$$

Combining the above inequality with (23), we obtain

$$\frac{\alpha_k}{2\eta} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{I}, \frac{\alpha_{\max}}{\eta}) \leq F(\mathbf{x}_k) - F(\mathbf{x}_{k+1}). \quad (28)$$

By using (17), we get

$$\frac{\nu \alpha_{\min}}{\eta} (F(\mathbf{x}_k) - F^*) \leq F(\mathbf{x}_k) - F^* - (F(\mathbf{x}_{k+1}) - F^*).$$

Rearranging the above inequality, we reach

$$F(\mathbf{x}_{k+1}) - F^* \leq \left(1 - \frac{\nu \alpha_{\min}}{\eta}\right) (F(\mathbf{x}_k) - F^*). \quad (29)$$

By letting $\alpha_k \leq \min\left\{\frac{\eta}{\nu}, \frac{\eta}{L}\right\}$, we have $(1 - \frac{\nu \alpha_{\min}}{\eta}) > 0$. Then applying (29) recursively, we obtain

$$F(\mathbf{x}_{k+1}) - F^* \leq \left(1 - \frac{\nu \alpha_{\min}}{\eta}\right)^k (F(\mathbf{x}_1) - F^*).$$

Now we show another formulation of convergence by uniformly sampling the output. Summing up (23) from $k = 1$ to K , we get

$$\begin{aligned}
\sum_{k=1}^K \frac{\alpha_k}{2} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) &\leq F(\mathbf{x}_1) - F(\mathbf{x}_K) \\
&\leq F(\mathbf{x}_1) - F^* \quad (30)
\end{aligned}$$

By uniformly sampling one of the previous iterates at $K - 1$ th iteration as the output $\mathbf{x}_{k'}$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}_h^C(\mathbf{x}_{k'}, \nabla f(\mathbf{x}_{k'}), \mathbf{I}, \frac{\alpha_{k'}}{\eta})\right] &= \frac{\sum_{k=1}^K \frac{\mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{I}, \frac{\alpha_k}{\eta})}{K}}{\sum_{k=1}^K \frac{\eta \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k)}{K}}, \quad (31) \\
&\leq \frac{\sum_{k=1}^K \frac{\mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k)}{K}}{\sum_{k=1}^K \frac{\eta \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k)}{K}},
\end{aligned}$$

where the inequality comes from (27). Summing up (23) from $k = 1$ to K , we obtain

$$\begin{aligned}
\sum_{k=1}^K \frac{\alpha_k}{2} \mathcal{D}_h^C(\mathbf{x}_k, \nabla f(\mathbf{x}_k), \mathbf{B}_k, \alpha_k) &\leq F(\mathbf{x}_1) - F(\mathbf{x}_{K+1}) \\
&\leq F(\mathbf{x}_1) - F^*.
\end{aligned}$$

Together with (31), we get

$$\mathbb{E}\left[\mathcal{D}_h^C(\mathbf{x}_{k'}, \nabla f(\mathbf{x}_{k'}), \mathbf{I}, \frac{\alpha_{k'}}{\eta})\right] \leq \frac{2\eta(F(\mathbf{x}_1) - F^*)}{\alpha_{\min} K}.$$

By invoking (17), we get the desired result

$$\mathbb{E}[F(\mathbf{x}_{k'}) - F^*] \leq \frac{\eta(F(\mathbf{x}_1) - F^*)}{\nu \alpha_{\min} K}.$$

Clearly, $F(\mathbf{x}_{k'})$ converges to F^* in expectation as $K \rightarrow \infty$.

REFERENCES

- [1] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging*. SPIE Optical Engineering Press Bellingham, 2000.
- [2] M. A. Brown and R. C. Semelka, *MRI: Basic Principles and Applications*. John Wiley & Sons, 2011.
- [3] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- [4] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nitka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions

- (GRAPPA),” *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.
- [5] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
 - [6] M. Guerquin-Kern, M. Haberland, K. P. Pruessmann, and M. Unser, “A fast wavelet-based reconstruction method for magnetic resonance imaging,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.
 - [7] M. V. Zibetti, E. S. Helou, R. R. Regatte, and G. T. Herman, “Monotone FISTA with variable acceleration for compressed sensing magnetic resonance imaging,” *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 109–119, 2018.
 - [8] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
 - [9] T. Hong, L. Hernandez-Garcia, and J. A. Fessler, “A complex quasi-Newton proximal method for image reconstruction in compressed sensing MRI,” *IEEE Transactions on Computational Imaging*, vol. 10, pp. 372 – 384, 2024.
 - [10] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
 - [11] S. Ravishanker and Y. Bresler, “MR image reconstruction from highly undersampled k-space data by dictionary learning,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.
 - [12] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, “Compressive sensing via nonlocal low-rank regularization,” *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.
 - [13] J. A. Fessler, “Model-based image reconstruction for MRI,” *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 81–9, Jul. 2010.
 - [14] —, “Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms,” *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 33–40, 2020.
 - [15] R. Heckel, M. Jacob, A. Chaudhari, O. Perlman, and E. Shimron, “Deep learning for accelerated and robust MRI reconstruction,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 37, no. 3, pp. 335–368, 2024.
 - [16] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, “Accelerating magnetic resonance imaging via deep learning,” in *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 514–517.
 - [17] H. K. Aggarwal, M. P. Mani, and M. Jacob, “MoDL: Model-based deep learning architecture for inverse problems,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, 2018.
 - [18] D. Gilton, G. Ongie, and R. Willett, “Deep equilibrium architectures for inverse problems in imaging,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1123–1133, 2021.
 - [19] Z. Ramzi, G. Chaithya, J.-L. Starck, and P. Ciuciu, “NC-PDNet: A density-compensated unrolled network for 2D and 3D non-Cartesian MRI reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1625–1638, 2022.
 - [20] Z. Wang, X. Yu, C. Wang, W. Chen, J. Wang, Y.-H. Chu, H. Sun, R. Li, P. Li, F. Yang *et al.*, “One for multiple: Physics-informed synthetic data boosts generalizable deep learning for fast MRI reconstruction,” *arXiv preprint arXiv:2307.13220*, 2023.
 - [21] B. Shi, Y. Wang, and D. Li, “Provable general bounded denoisers for snapshot compressive imaging with convergence guarantee,” *IEEE Transactions on Computational Imaging*, vol. 9, pp. 55–69, 2023.
 - [22] Y. Song, L. Shen, L. Xing, and S. Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *International Conference on Learning Representations*, 2021.
 - [23] H. Chung and J. C. Ye, “Score-based diffusion models for accelerated MRI,” *Medical Image Analysis*, vol. 80, p. 102479, 2022.
 - [24] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in *IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.
 - [25] Y. Romano, M. Elad, and P. Milanfar, “The little engine that could: Regularization by denoising (RED),” *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.
 - [26] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
 - [27] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
 - [28] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, “Plug-and-play priors for bright field electron tomography and sparse interpolation,” *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 408–423, 2016.
 - [29] S. Ono, “Primal-dual plug-and-play image restoration,” *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1108–1112, 2017.
 - [30] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, “Learning proximal operators: Using denoising networks for regularizing inverse imaging problems,” in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, Oct. 2017, pp. 1799–1808.
 - [31] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, “Plug-and-Play unplugged: Optimization free reconstruction using consensus equilibrium,” *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2001–2020,

- 2018.
- [32] B. Shi, Q. Lian, and H. Chang, “Deep prior-based sparse representation model for diffraction imaging: A plug-and-play method,” *Signal Processing*, vol. 168, p. 107350, 2020.
 - [33] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, “Plug-and-play image restoration with deep denoiser prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6360–6376, 2021.
 - [34] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, “Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery,” *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 105–116, 2020.
 - [35] N. Parikh, S. Boyd *et al.*, “Proximal algorithms,” *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
 - [36] M. L. Pendu and C. Guillemot, “Preconditioned plug-and-play ADMM with locally adjustable denoiser for image restoration,” *SIAM Journal on Imaging Sciences*, vol. 16, no. 1, pp. 393–422, 2023.
 - [37] T. Hong, X. Xu, J. Hu, and J. A. Fessler, “Provable preconditioned plug-and-play approach for compressed sensing MRI reconstruction,” *IEEE Transactions on Computational Imaging*, vol. 10, pp. 372 – 384, 2024.
 - [38] T. Hong, Y. Romano, and M. Elad, “Acceleration of RED via vector extrapolation,” *Journal of Visual Communication and Image Representation*, p. 102575, 2019.
 - [39] E. T. Reehorst and P. Schniter, “Regularization by denoising: Clarifications and new interpretations,” *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 52–67, 2018.
 - [40] T. Hong, I. Yavneh, and M. Zibulevsky, “Solving RED with weighted proximal methods,” *IEEE Signal Processing Letters*, vol. 27, pp. 501–505, 2020.
 - [41] S. H. Chan, X. Wang, and O. A. Elgendy, “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.
 - [42] Y. Sun, B. Wohlberg, and U. S. Kamilov, “An online plug-and-play algorithm for regularized image reconstruction,” *IEEE Transactions on Computational Imaging*, vol. 5, no. 3, pp. 395–408, Sep. 2019.
 - [43] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, “Plug-and-play methods provably converge with properly trained denoisers,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 5546–5557.
 - [44] X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov, “Provable convergence of plug-and-play priors with MMSE denoisers,” *IEEE Signal Processing Letters*, vol. 27, pp. 1280–1284, 2020.
 - [45] B. Shi and D. Li, “Provably bounded dynamic sparsifying transform network for compressive imaging,” *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
 - [46] R. G. Gavaskar, C. D. Athalye, and K. N. Chaudhury, “On plug-and-play regularization using linear denoisers,” *IEEE Transactions on Image Processing*, vol. 30, pp. 4802–4813, 2021.
 - [47] Y. Sun, Z. Wu, X. Xu, B. Wohlberg, and U. S. Kamilov, “Scalable plug-and-play ADMM with convergence guarantees,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 849–863, 2021.
 - [48] J. Liu, S. Asif, B. Wohlberg, and U. Kamilov, “Recovery analysis for plug-and-play priors using the restricted eigenvalue condition,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 5921–5933, 2021.
 - [49] R. Cohen, M. Elad, and P. Milanfar, “Regularization by denoising via fixed-point projection (RED-PRO),” *SIAM Journal on Imaging Sciences*, vol. 14, no. 3, pp. 1374–1406, 2021.
 - [50] 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,” *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, 2023.
 - [51] R. Gribonval and M. Nikolova, “A characterization of proximity operators,” *Journal of Mathematical Imaging and Vision*, vol. 62, no. 6, pp. 773–789, 2020.
 - [52] —, “On Bayesian estimation and proximity operators,” *Applied and Computational Harmonic Analysis*, vol. 50, pp. 49–72, 2021.
 - [53] M. Terris, A. Repetti, J.-C. Pesquet, and Y. Wiaux, “Building firmly nonexpansive convolutional neural networks,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8658–8662.
 - [54] A. Pramanik, M. B. Zimmerman, and M. Jacob, “Memory-efficient model-based deep learning with convergence and robustness guarantees,” *IEEE Transactions on Computational Imaging*, vol. 9, pp. 260–275, 2023.
 - [55] S. Aziznejad, H. Gupta, J. Campos, and M. Unser, “Deep neural networks with trainable activations and controlled Lipschitz constant,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 4688–4699, 2020.
 - [56] K. Bredies, J. Chirinos-Rodriguez, and E. Naldi, “Learning firmly nonexpansive operators,” *arXiv preprint arXiv:2407.14156*, 2024.
 - [57] R. Cohen, Y. Blau, D. Freedman, and E. Rivlin, “It has potential: Gradient-driven denoisers for convergent solutions to inverse problems,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 18 152–18 164, 2021.
 - [58] S. Hurault, A. Leclaire, and N. Papadakis, “Gradient step denoiser for convergent plug-and-play,” *arXiv preprint arXiv:2110.03220*, 2021.
 - [59] Z. Fang, S. Buchanan, and J. Sulam, “What’s in a prior? learned proximal networks for inverse problems,” *arXiv preprint arXiv:2310.14344*, 2023.
 - [60] S. Chaudhari, S. Pranav, and J. M. Moura, “Gradient networks,” *IEEE Transactions on Signal Processing*, vol. 73, pp. 324 – 339, 2024.
 - [61] H. Y. Tan, S. Mukherjee, J. Tang, and C.-B. Schönlieb, “Provably convergent plug-and-play quasi-Newton methods,” *SIAM Journal on Imaging Sciences*, vol. 17, no. 2,

- pp. 785–819, 2024.
- [62] L. Stella, A. Themelis, and P. Patrinos, “Forward–backward quasi-Newton methods for nonsmooth optimization problems,” *Computational Optimization and Applications*, vol. 67, no. 3, pp. 443–487, 2017.
 - [63] M. Schmidt, E. Berg, M. Friedlander, and K. Murphy, “Optimizing costly functions with simple constraints: A limited-memory projected quasi-Newton algorithm,” in *Artificial Intelligence and Statistics*. PMLR, 2009, pp. 456–463.
 - [64] J. D. Lee, Y. Sun, and M. A. Saunders, “Proximal Newton-type methods for minimizing composite functions,” *SIAM Journal on Optimization*, vol. 24, no. 3, pp. 1420–1443, 2014.
 - [65] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, “Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function,” *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 107–132, 2014.
 - [66] S. Bonettini, I. Loris, F. Porta, and M. Prato, “Variable metric inexact line-search-based methods for nonsmooth optimization,” *SIAM Journal on Optimization*, vol. 26, no. 2, pp. 891–921, 2016.
 - [67] S. Karimi and S. Vavasis, “IMRO: A proximal quasi-Newton method for solving ℓ_1 -regularized least squares problems,” *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 583–615, 2017.
 - [68] S. Becker, J. Fadili, and P. Ochs, “On quasi-Newton forward-backward splitting: Proximal calculus and convergence,” *SIAM Journal on Optimization*, vol. 29, no. 4, pp. 2445–2481, 2019.
 - [69] A. Repetti and Y. Wiaux, “Variable metric forward-backward algorithm for composite minimization problems,” *SIAM Journal on Optimization*, vol. 31, no. 2, pp. 1215–1241, 2021.
 - [70] M. Zhang and S. Li, “A proximal stochastic quasi-Newton algorithm with dynamical sampling and stochastic line search,” *Journal of Scientific Computing*, vol. 102, no. 1, pp. 1–36, 2025.
 - [71] M. Osborne and L. Sun, “A new approach to symmetric rank-one updating,” *IMA Journal of Numerical Analysis*, vol. 19, no. 4, pp. 497–507, 1999.
 - [72] F. Curtis, “A self-correcting variable-metric algorithm for stochastic optimization,” in *International Conference on Machine Learning*. PMLR, 2016, pp. 632–641.
 - [73] X. Wang, X. Wang, and Y.-X. Yuan, “Stochastic proximal quasi-Newton methods for non-convex composite optimization,” *Optimization Methods and Software*, vol. 34, no. 5, pp. 922–948, 2019.
 - [74] Y. Nesterov, *Lectures on Convex Optimization*. Springer, 2018.
 - [75] T. Hong and I. Yavneh, “On adapting Nesterov’s scheme to accelerate iterative methods for linear problems,” *Numerical Linear Algebra with Applications*, vol. 29, no. 2, p. e2417, 2022.
 - [76] H. Karimi, J. Nutini, and M. Schmidt, “Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition,” in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19–23, 2016, Proceedings, Part I 16*. Springer, 2016, pp. 795–811.
 - [77] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, “Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality,” *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.
 - [78] T. Hong, U. Villa, and J. A. Fessler, “A convergent generalized Krylov subspace method for compressed sensing mri reconstruction with gradient-driven denoisers,” *arXiv preprint arXiv:2508.11219*, 2025.
 - [79] J. E. Dennis, Jr and J. J. Moré, “Quasi-Newton methods, motivation and theory,” *SIAM Review*, vol. 19, no. 1, pp. 46–89, 1977.
 - [80] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno *et al.*, “fastMRI: An open dataset and benchmarks for accelerated MRI,” *arXiv preprint arXiv:1811.08839*, 2018.
 - [81] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA,” *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014.
 - [82] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
 - [83] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “Coil compression for accelerated imaging with Cartesian sampling,” *Magnetic Resonance in Medicine*, vol. 69, no. 2, pp. 571–582, 2013.
 - [84] H. Li and Z. Lin, “Accelerated proximal gradient methods for nonconvex programming,” *Advances in Neural Information Processing Systems*, vol. 28, 2015.
 - [85] S. Hurault, A. Leclaire, and N. Papadakis, “Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization,” in *International Conference on Machine Learning*. PMLR, 2022, pp. 9483–9505.
 - [86] B. Amos, L. Xu, and J. Z. Kolter, “Input convex neural networks,” in *International Conference on Machine Learning*. PMLR, 2017, pp. 146–155.