

1 **An Unrolled Implicit Regularization Network for Joint Image and Sensitivity
2 Estimation in Parallel MR Imaging with Convergence Guarantee***

3 Yan Yang[†], Yizhou Wang[‡], Jiazen Wang[†], Jian Sun[§], and Zongben Xu[†]

5 **Abstract.** Parallel imaging (PI), relying on multi-coils to sense k -space data, is an effective technique to ac-
6 celerate magnetic resonance imaging by exploiting spatial sensitivity coding of multiple coils, with
7 an integrated compressive sensing (CS) technology to achieve higher acceleration. In this paper, we
8 propose a novel non-convex reconstruction model and its proximal alternating linearized minimiza-
9 tion (PALM) algorithm for PI in a blind setting that MR image and multi-channel sensitivity maps
10 are jointly estimated, regularized by image and sensitivity regularizers. Instead of hand-crafting the
11 image and sensitivity regularizers, we propose to unroll the PALM algorithm to be a deep network
12 for Blind Parallel MRI, dubbed as BPMRI-Net, with two learnable sub-networks to substitute the
13 proximal operators of the image and sensitivity regularizers. We theoretically prove the linear con-
14 vergence of BPMRI-Net as an iterative algorithm, which alternately updates two variables based
15 on the learnable proximal operators. The learned BPMRI-Net can simultaneously output the MR
16 image and sensitivity maps from under-sampled multi-channel k -space data even when the number
17 of low-frequency sampling lines in the center of k -space is small. Numerical results demonstrate the
18 effectiveness of our method with state-of-the-art reconstruction accuracy.

19 **Key words.** Parallel MR imaging, compressive sensing, proximal operator learning, convergence analysis.

20 **AMS subject classifications.** 65K10, 94A08, 65J22, 97R40, 68Q32, 65J20

21 **1. Introduction.** Magnetic resonance imaging (MRI) is a non-invasive medical imaging
22 technology providing clear contrast of different soft tissues to depict the internal anatomical
23 and functional information of the body for diagnosis. However, due to the physics of MR
24 imaging, one limitation of MRI is its long data acquisition time, causing discomfort for patients
25 or introducing motion-related artifacts in MR images. This limitation restricts the clinical
26 applications in dynamic imaging or fetal imaging.

27 To accelerate MRI, parallel imaging (PI) and compressive sensing (CS) are two popular
28 techniques in MRI. The parallel imaging reduces the number of k -space data samples for
29 acceleration by spatial sensitivity coding of multiple coils [19, 43]. Compressive sensing MRI
30 (CS-MRI) is based on sparse sampling in the k -space at a low sampling rate and reconstructs
31 MR images based on the image sparsity [33, 34]. Combining parallel imaging with compressive
32 sensing, i.e., CS-PI, has been a fundamental acceleration technique in MRI [7, 22, 31]. CS-PI
33 is based on the following data acquisition mechanism, i.e., a linear forward imaging model for

*The first two authors contributed equally. Part of this work was published in MICCAI (2019) [37].

Funding: This work was supported by the National Key R&D Program 2022YFA1004201 and the NSFC with grant No. 12090020, 12090021, 12125104, 12026605, U20B2075.

[†]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P. R. China (yangyan@xjtu.edu.cn, jzwang@stu.xjtu.edu.cn, zbxu@xjtu.edu.cn).

[‡]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, United States (wyzjack990122@gmail.com).

[§]Corresponding author. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P. R. China (jiansun@xjtu.edu.cn).

34 each coil indexed by l :

35 (1.1)
$$\mathbf{y}_l = \mathbf{M}\mathcal{F}(\mathbf{s}_l \odot \mathbf{u}) + \mathbf{e}_l \quad l \in \{1, 2, \dots, N_c\}.$$

36 It relates the under-sampling k -space data of l -th coil $\mathbf{y}_l \in \mathbb{C}^M$ to the sensitivity map $\mathbf{s}_l \in \mathbb{C}^N$
 37 and the underlying MR image to reconstruct $\mathbf{u} \in \mathbb{C}^N$ ($M \ll N$). $\mathbf{e}_l \in \mathbb{C}^M$ denotes a complex-
 38 valued Gaussian noise in l -th coil and \odot is a Hadamard product. The number of coils is N_c .
 39 $\mathcal{F} \in \mathbb{C}^{N \times N}$ is a Fourier transform matrix and $\mathbf{M} \in \mathbb{C}^{M \times N}$ is a sampling matrix in k -space.
 40 CS-PI aims to reconstruct high-quality MR images from limited k -space data observed by
 41 multiple coils. The measurements of each coil can reconstruct a part of the clear MR image,
 42 and its sensitivity map determines its observable range.

43 According to the role of coil sensitivities in the reconstruction method, there are three
 44 categories of methods for CS-PI, including **explicit-calibration**, **auto-calibration** and
 45 **calibration-less** methods. The explicit-calibration methods exploit the coil sensitivity maps
 46 explicitly which are estimated by pre-scanning calibration images [7, 31, 43]. Based on the
 47 estimated sensitivity maps (i.e., $\mathbf{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_l, \dots, \mathbf{s}_{N_c}\}$), most of these methods solve an
 48 image reconstruction model w.r.t. the image \mathbf{u} :

49 (1.2)
$$\underset{\mathbf{u}}{\text{minimize}} \sum_{l=1}^{N_c} \|\mathbf{M}\mathcal{F}(\mathbf{s}_l \odot \mathbf{u}) - \mathbf{y}_l\|_2^2 + \lambda P(\mathbf{u}),$$

50 where $P(\cdot)$ is a predefined regularization function for modeling the MR image prior and λ
 51 is a preset regularization parameter. For example, the function $P(\cdot)$ can be taken as To-
 52 tal Variation (TV) or Total Generalized Variation (TGV) seminorm [7, 12, 31], non-local
 53 sparsity [30, 49], sparsity in wavelets domain [29], and dictionary or transform learning spar-
 54 sity [3, 58] regularizations. This kind of approach allows optimal reconstruction in the prin-
 55 ciple of minimum mean square error [56] and can be applied to different k -space sampling
 56 patterns. However, possible misalignment between calibration scans and subsequent accel-
 57 erated imaging may lead to sensitivity miscalibration errors, which deteriorates the quality
 58 of the reconstructed MR images and introduces visible artifacts. In order to alleviate this
 59 problem, the auto-calibration methods extract auto-calibration signals (ACS) from a fully
 60 sampled central region of k -space to measure the sensitivity maps or estimate correspond-
 61 ing interpolation kernels in k -space to implicitly use sensitivity information. Representative
 62 methods are Auto-SMASH [25], GRAPPA [19], SPIRiT [35], ESPIRiT [56], PRUNO [69],
 63 AC-LORAKS [20], etc. For example, GRAPPA exploits the correlation from ACS to esti-
 64 mate a translation variant interpolation kernel, and then the non-acquired k -space value is
 65 synthesized by acquired neighboring k -space data using this interpolation kernel. SPIRiT
 66 solves an inverse problem with two separate data consistency constraints, including consis-
 67 tency with auto-calibration data and consistency with acquisition data, which can also be
 68 flexibly combined with additional image priors. AC-LORAKS uses ACS to estimate the null
 69 spaces associated with the low-rank matrices and speeds up the computational efficiency of
 70 matrix completion. The reconstruction accuracy of these methods depends heavily on the
 71 accuracy of the estimated sensitivity profile, especially in the case that a limited number of
 72 ACS lines exist at a high acceleration rate.

The calibration-less methods do not require an extra calibration stage. This type of method may directly reconstruct multi-channel full k -space data or multi-channel sensitivity encoded images (i.e., $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l, \dots, \mathbf{v}_{N_c} | \mathbf{v}_l = \mathbf{s}_l \odot \mathbf{u}, l \in \{1, 2, \dots, N_c\}\}$), and synthesizes the final MR image by the square root of the sum of squares (SoS) of the reconstructed multi-coil sensitivity encoded images (i.e., $\mathbf{u} = \sqrt{\sum_{l=1}^{N_c} \mathbf{v}_l^* \odot \mathbf{v}_l}$). Another idea is to solve an optimization problem that takes sensitivity as a latent variable to be optimized in an iterative reconstruction algorithm. By exploiting the relationship of multi-channel k -space data, techniques such as CLEAR [54], ALOHA [26], and P-LORAKS [21, 22] solve a constrained optimization problem with structured low-rank matrix modeling. For example, P-LORAKS designs a structured low-rank matrix that can simultaneously capture support, phase, and parallel imaging constraints. However, high computational complexity is the main disadvantage of these structured low-rank matrix completion methods. Some methods [10, 36, 57] bypass the sensitivity information and reconstruct multi-channel sensitivity encoded images by solving an optimization problem penalized by a predefined regularization function on the prior of sensitivity encoded images. Since multi-channel images share similar supports, joint sparsity priors are often used to mine the redundancy of cross-channel images. The joint sparsity representation can be based on fixed TV [10], wavelet transform [10, 36] or adaptive learning by image patches [57]. Although these methods are flexible and robust to noise, they do not utilize sensitivity information and combine real-valued MR images with the SoS operation, resulting in the loss of phase information, which is important for clinical applications such as flow imaging and phase-contrast MRI [59], etc. Moreover, in the category of calibration-less methods, methods such as NLINV [55], ENLIVE [24], JSENSE [67], and Sparse BLIP [48], treat PI as a blind inverse problem, which jointly estimates the MR image \mathbf{u} and sensitivity maps \mathbf{S} . They use an iterative algorithm to optimize the following reconstruction model regularized by both MR image and sensitivity priors:

$$(1.3) \quad \underset{\mathbf{u}, \mathbf{S}}{\text{minimize}} \sum_{l=1}^{N_c} \|\mathbf{M}\mathcal{F}(\mathbf{s}_l \odot \mathbf{u}) - \mathbf{y}_l\|_2^2 + \lambda P(\mathbf{u}) + \gamma R(\mathbf{S}),$$

where the regularization term $P(\cdot)$ associated with image prior is often used to describe the sparsity of MR image [27], and the sensitivity prior $R(\cdot)$ can be taken as TV of the image domain [48] for sparseness, polynomial regularization [67] or Tikhonov regularization of an invertible weighting transform domain [24] to penalize high frequencies in the coil profiles. These methods correct the estimated sensitivity maps during iterations and can improve the performance of PI reconstruction. They are inherently non-convex and usually solved by alternate iterations, which require more update steps and longer running time. For the above model-based methods, it is challenging to choose optimal regularization models of MR image and sensitivity (i.e., $P(\cdot)$ and $R(\cdot)$), or determine the optimal hyper-parameters in models and optimization algorithms, such as λ , γ , and step sizes of variable update, either of which can affect the accuracy and convergence speed of the reconstruction algorithms.

Recently, deep learning has become an effective tool for solving inverse problems and has also been successfully applied to CS-PI. The basic idea is to use a general deep neural network with a large number of parameters to directly fit a mapping from under-sampled k -space data of multi-channel to aliasing-free multi-channel sensitivity encoded images or full k -space

114 data [28, 71]. For instance, [71] and [39] rely on a generative adversarial network (GAN) and
 115 an optimal transport driven cycle-consistent GAN (OT-cycleGAN) to learn this mapping.
 116 However, such networks are purely data-driven that do not incorporate the domain knowledge
 117 of the specific problem at hand. To incorporate the imaging mechanism of MRI, Wang et
 118 al. [14] and Schlemper et al. [46] further introduce the imaging model or data consistency in
 119 k -space into the network structure. Learning VN [23] and MoDL [1] are two explicit-calibration
 120 CS-PI methods, which learn a Fields of Experts (FOE) model or a CNN-based denoiser by
 121 unfolding the (conjugate) gradient descent procedure into a deep network. The sensitivity
 122 maps embedded in these networks are estimated by a calibration step using ESPIRiT [56].
 123 These methods achieved promising results and high reconstruction speed, however, they rely
 124 on estimating the sensitivities of coils with a separate step. Furthermore, the networks lack
 125 theoretical analysis which is challenging for neural network-based reconstruction models.

126 This work proposed an iterative deep learning-based method for CS-PI with theoretical
 127 convergence analysis, being able to jointly estimate MR images and the coil sensitivities with
 128 high acceleration factors or fewer ACS lines in k -space. Specifically, based on the imaging
 129 mechanism of PI, we first formulate the reconstruction problem as a non-convex optimization
 130 model for joint estimation of MR image and sensitivity maps with PI imaging model as data
 131 term, regularized by two pre-defined regularizers and two undetermined image and sensitivity
 132 priors. Then we use an iterative algorithm based on proximal alternating linearized minimiza-
 133 tion (PALM) to alternately update the reconstructed MR image and sensitivity maps. Next,
 134 we propose to unroll the PALM algorithm to be a novel non-conventional deep architecture,
 135 dubbed as **BPMRI-Net**, which can simultaneously update a complex-valued MR image and
 136 sensitivity maps of multiple coils from under-sampled multi-channel k -space data to iteratively
 137 correct the estimations during the reconstruction process. Instead of setting the regularization
 138 terms for priors of image and sensitivity by hand, we design two deep convolutional networks
 139 as sub-networks to learn proximal operators in the PALM iterations to implicitly learn their
 140 corresponding regularizers. All undetermined parameters in the reconstruction model, itera-
 141 tive algorithm and sub-networks are transferred to be the parameters of the proposed network
 142 BPMRI-Net, which can be learned by end-to-end learning based on fully sampled data. In the
 143 testing phase, BPMRI-Net can perform a joint reconstruction of MR image and sensitivities.
 144 In addition, since the network structure is derived from PALM iterations, we theoretically
 145 prove the convergence of the proposed BPMRI-Net as an iterative algorithm.

146 **1.1. Our contributions.** The main contributions of this paper are summarized as follows.

- 147 (1) We propose a novel non-convex reconstruction model for PI in a blind setting with
 148 regularization on both MR image and sensitivity priors, and its PALM optimization
 149 algorithm, which can jointly estimate the MR image and multi-channel sensitivity
 150 maps. Furthermore, we unroll the iterations of the PALM algorithm into a novel
 151 deep architecture, in which the proximal operators were learned using designed deep
 152 convolutional neural networks to implicitly learn the corresponding regularizers.
- 153 (2) By integrating the Lipschitz properties of objective function and sub-networks, we
 154 theoretically prove that the proposed BPMRI-Net as an iterative algorithm with two
 155 learnable proximal operators can linearly converge to a fixed point asymptotically.
- 156 (3) After training, the MR image and sensitivity maps are gradually updated with the net-

work iterations, and finally, a complex-valued MR image is reconstructed. Compared with the traditional model-based PI methods and existing deep learning methods, the proposed method achieved state-of-the-art reconstruction accuracy, especially in the case of fewer ACS lines. On average, BPMRI-Net outperforms traditional methods of AC-LORAKS, Fast-JTV, and LINDBERG by at least 2 dB and outperforms deep learning methods of DC-CNN, Learning VN, and MoDL by at least 0.5 dB, under 4, 6, and 8 times 1D Cartesian random under-sampling on a knee dataset.

The preliminary version of this paper was published in the conference version [37], to our best knowledge, which is the first work for CS-PI reconstruction in a blind setting that was modeled by an unrolled deep network. Compared with the preliminary version [37], this paper further extends the conference version in two aspects. On the one hand, we reformulate the reconstruction model with mixed pre-defined regularizers and undetermined regularization functions on MR image and sensitivity maps and use PALM to update the MR images and sensitivity maps alternately. In the conference version, the reconstruction model contains two purely undetermined regularization functions and the Half-quadratic Splitting (HQS) optimization algorithm is adopted to alternately update the three variables including multi-channel sensitivity encoded images, sensitivity maps and MR image. We further theoretically prove the convergence of the proposed network as an iterative algorithm in this work. On the second hand, the revised model and algorithm are justified to perform better than the BPMRI-Net proposed in [37], and achieved state-of-the-art results in experiments.

1.2. Related work. Model-driven deep learning method is an effective approach to design the network structure by unfolding a truncated iterative process into a hierarchical architecture or inserting a deep network into the iterations of optimization algorithm [63, 70]. It combines the advantages of the model-based method and the deep learning method to make network performance better explainable compared with the black-box deep network. It was first applied to learning sparse coding (LISTA) [18] and Markov random field [51]. Then it was successfully used to solve inverse problems and greatly improved the accuracy of reconstruction. For example, [47, 70] are used for image restoration, ADMM-CSNet [66] and ISTA-Net [68] are applied to CS and [1, 23, 46, 65] are for medical imaging.

Recently, there have been some attempts to theoretically analyze the convergence of model-driven deep learning algorithms to understand their efficient approximation. [61] proposes that learning the unfolded iterative hard thresholding (IHT) algorithm may improve the restricted isometry constant (RIC) of the dictionary to speed up the convergence. In [11], the linear convergence of unfolded ISTA to the target sparse signal is justified under the introduced weight structure, which is superior to the sub-linear convergence of the original ISTA algorithm. Furthermore, using this coupled weight structure, a deep neural network named LePOM is proposed in [64] based on unfolding the proximal operator algorithm for solving a non-convex sparsity recovery problem, which improves the performance and relieves the computational burden of the algorithm. The output of LePOM is proved to converge linearly early under certain conditions. [16] attempts to relate LISTA to an inexact projected gradient descent (IPGD) method to trade-off between approximation error and convergence rate. A Linearized ADMM algorithm inspired network (D-LADMM) [60] is proposed to solve linearly constrained problems. The authors rigorously prove the existence of a set of parameters for

200 the global convergence of D-LADMM and show that this set of parameters can be obtained
 201 by proper training of D-LADMM. The above methods make use of the analytic form of a
 202 proximal operator, the optimality condition of the optimization problem corresponding to the
 203 proximal operator or the sparsity of signals for theoretical analysis. In particular, they are
 204 aimed at learning specific regularization functions, such as ℓ_0 norm, ℓ_1 norm, and minimax
 205 concave penalty (MCP), etc., resulting in limited learning ability. Our proposed network
 206 directly learns the proximal operators using the deep convolutional networks to model the
 207 nonlinear mappings and learns the regularization functions implicitly. Thus it is difficult to
 208 use the analytical properties of the proximal operator to analyze the theoretical properties of
 209 the proposed network. In addition, we prove the convergence of the proposed network without
 210 assumption on the signal sparsity.

211 The plug-and-play deep learning method introduces an operator regularization by applying
 212 a pre-trained deep neural network as a denoiser with additive white Gaussian noise, which has
 213 been a widely applicable method for solving inverse problems [2, 44, 52]. According to the fixed
 214 point theory, [40, 52] theoretically establish the convergence of the proposed algorithms under
 215 a firmly non-expansive condition on the denoisers. To plug in more general deep networks,
 216 Ryu et al. [45] proposed PNP-FBS and PNP-ADMM algorithms without decreasing step
 217 size, in which the denoiser satisfies a certain Lipschitz condition to ensure convergence. The
 218 current PNP algorithms deal with the optimization problem with univariate updates and
 219 convex data fidelity terms. The blind inverse problem with two regularization terms has
 220 not been solved. Furthermore, pre-trained deep networks lack adaptability to different fields
 221 of data and different noise types and noise levels, limiting the accuracy of the algorithms.
 222 In contrast to these methods, our network is derived from a truncated iterative algorithm
 223 optimizing a blind non-convex model with two undetermined regularization terms, in which
 224 two variables are updated alternately. Our method is based on end-to-end training that the
 225 proximal operators are trained within the iterative optimization algorithm, different from the
 226 plug-and-play method that inserts a pre-trained denoiser into an iterative algorithm.

227 Estimating sensitivities in parallel MRI using the deep learning approach was first pro-
 228 posed in [37] and [15]. Recently, E2-VarNet [50] estimates the sensitivity maps using a U-Net
 229 based on a variational network. The Deep J-Sense method is designed in [4], which combines
 230 MoDL and J-Sense. It jointly estimates the image and sensitivity kernel in k -space, by unfold-
 231 ing an alternating optimization algorithm. The methods in [32, 8] use a deep neural network
 232 to learn the image prior and combine it with the NLINV method [55] for joint estimation of
 233 coil sensitivities and image content, where the regularization terms over the sensitivity maps
 234 are predefined without learning. Compared with the methods in [15] and E2E-VarNet, the
 235 proposed BPMRI-Net iteratively updates the sensitivity maps with the network iterations,
 236 together with the updated MR images. Different from Deep J-Sense, our proposed method
 237 unrolls a different iterative algorithm with a convergence guarantee and estimates the sensi-
 238 tivity maps in the image domain instead of the sensitivity kernels in k -space.

239 **1.3. Organization of the paper.** The remainder of this paper is organized as follows.
 240 In section 2, we propose the reconstruction model, optimization algorithm, and deep net-
 241 work for CS-PI. We prove convergence of the proposed BPMRI-Net as an iterative algorithm
 242 in section 3. Numerical experiments are conducted to verify the effectiveness of our method

243 in section 4. Conclusion and discussion are in section 5.

244 **2. Proposed method.** In this section, we first propose a reconstruction model for parallel
245 MR imaging. Then we design an iterative algorithm to solve the model. We further derive
246 our deep network by unrolling this iterative algorithm.

247 **2.1. Model.** Assuming that we are given partial k -space data of N_c channels captured
248 by multi-receivers $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_l, \dots, \mathbf{y}_{N_c}\}$, we aim to estimate both the reconstructed
249 MR image \mathbf{u} and the sensitivity maps $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_l, \dots, \mathbf{s}_{N_c}\}$. Based on the imaging
250 mechanism of parallel MR, we propose the following reconstruction model:

$$251 \quad (2.1) \quad \underset{\mathbf{u}, \mathbf{S}}{\text{minimize}} \quad E(\mathbf{u}, \mathbf{S}) := H(\mathbf{u}, \mathbf{S}) + \lambda P(\mathbf{u}) + \gamma \sum_{l=1}^{N_c} R(\mathbf{s}_l),$$

252 where

$$253 \quad (2.2) \quad H(\mathbf{u}, \mathbf{S}) = \frac{1}{2} \sum_{l=1}^{N_c} \|\mathbf{A}(\mathbf{s}_l \odot \mathbf{u}) - \mathbf{y}_l\|_2^2 + \frac{\alpha}{2} \|\mathbf{u}\|_2^2 + \frac{\beta}{2} \sum_{l=1}^{N_c} \|\mathbf{s}_l\|_2^2$$

254 is a smooth coupling function and $\mathbf{A} = \mathbf{M}\mathcal{F}$. The first term in $H(\mathbf{u}, \mathbf{S})$ enforces data con-
255 sistency among reconstructed MR image \mathbf{u} , estimated per-channel sensitivity map \mathbf{s}_l and its
256 under-sampled data \mathbf{y}_l in k -space. Minimizing this data term is a bilinear inverse problem,
257 which is ill-conditioned and usually does not guarantee the existence of a globally optimal
258 solution. $\|\mathbf{A}(\mathbf{s}_l \odot \mathbf{u}) - \mathbf{y}_l\|_2^2$ is a biconvex function, i.e., it is convex w.r.t. \mathbf{u} or \mathbf{s}_l if fixing
259 \mathbf{s}_l or \mathbf{u} , respectively. In $H(\mathbf{u}, \mathbf{S})$, we add squared ℓ_2 -regularization terms of the image and
260 sensitivity maps to transform the biconvex function to a strongly biconvex function, where
261 $\alpha > 0$ and $\beta > 0$ are corresponding regularization parameters. In order to further constrain
262 the solution space, $P(\cdot) : \mathbb{C}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ and $R(\cdot) : \mathbb{C}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ are two additional
263 undetermined regularization functions for characterizing the complex priors of the MR image
264 and the sensitivity maps respectively. They are assumed to be lower semi-continuous and
265 non-convex functions, which are not specified but implicitly learned from data. $\lambda > 0$ and
266 $\gamma > 0$ are regularization parameters corresponding to the relative weights of the undetermined
267 regularization terms of MR image and sensitivity maps respectively.

268 **2.2. Model optimization.** To optimize the unknown variables in Eqn. (2.1), we iteratively
269 update the variables \mathbf{u} and \mathbf{s}_l ($l \in \{1, 2, \dots, N_c\}$) by solving the following sub-problems:

$$270 \quad (2.3) \quad \begin{cases} \mathbf{u}^{(k+1)} \in \underset{\mathbf{u}}{\text{argmin}} \quad H(\mathbf{u}, \mathbf{S}^{(k)}) + \lambda P(\mathbf{u}), \\ \mathbf{s}_l^{(k+1)} \in \underset{\mathbf{s}_l}{\text{argmin}} \quad H(\mathbf{u}^{(k+1)}, \mathbf{S}) + \gamma R(\mathbf{s}_l), \quad (l \in \{1, 2, \dots, N_c\}), \end{cases}$$

271 where $k \in \{0, 1, \dots, K-1\}$, $\mathbf{S}^{(k)} = \{\mathbf{s}_1^{(k)}, \mathbf{s}_2^{(k)}, \dots, \mathbf{s}_l^{(k)}, \dots, \mathbf{s}_{N_c}^{(k)}\}$. By linearizing the function
272 H at points $\mathbf{u}^{(k)}$ and $\mathbf{s}_l^{(k)}$ ($l \in \{1, 2, \dots, N_c\}$) respectively, then the above sub-problems can
273 be transformed to

$$274 \quad (2.4) \quad \begin{cases} \mathbf{u}^{(k+1)} \in \underset{\mathbf{u}}{\text{argmin}} \langle \mathbf{u} - \mathbf{u}^{(k)}, \nabla_{\mathbf{u}} H(\mathbf{u}^{(k)}, \mathbf{S}^{(k)}) \rangle + \frac{\alpha}{2} \|\mathbf{u} - \mathbf{u}^{(k)}\|_2^2 + \lambda P(\mathbf{u}), \\ \mathbf{s}_l^{(k+1)} \in \underset{\mathbf{s}_l}{\text{argmin}} \langle \mathbf{s}_l - \mathbf{s}_l^{(k)}, \nabla_{\mathbf{s}_l} H(\mathbf{u}^{(k+1)}, \mathbf{S}^{(k)}) \rangle + \frac{\beta}{2} \|\mathbf{s}_l - \mathbf{s}_l^{(k)}\|_2^2 + \gamma R(\mathbf{s}_l), \end{cases}$$

275 where

276 (2.5)
$$\nabla_{\mathbf{u}} H(\mathbf{u}^{(k)}, \mathbf{S}^{(k)}) = \sum_{l=1}^{N_c} \{(\mathbf{s}_l^{(k)})^* \odot \mathbf{A}^\top [\mathbf{A}(\mathbf{s}_l^{(k)} \odot \mathbf{u}^{(k)}) - \mathbf{y}_l]\} + \alpha \mathbf{u}^{(k)},$$

277

278 (2.6)
$$\nabla_{\mathbf{s}_l} H(\mathbf{u}^{(k+1)}, \mathbf{S}^{(k)}) = (\mathbf{u}^{(k+1)})^* \odot \mathbf{A}^\top [\mathbf{A}(\mathbf{s}_l^{(k)} \odot \mathbf{u}^{(k+1)}) - \mathbf{y}_l] + \beta \mathbf{s}_l^{(k)},$$

279 are gradients of multivariate function H with respect to \mathbf{u} or \mathbf{s}_l . $a > 0$ and $d > 0$. The
280 superscripts * and \top represent conjugate and conjugate transpose operators, respectively.
281 $\langle \cdot, \cdot \rangle$ is a real-valued inner product of two complex vectors defined as $\langle \mathbf{x}, \mathbf{z} \rangle = \frac{1}{2}(\mathbf{x}^\top \mathbf{z} + \mathbf{z}^\top \mathbf{x})$.
282 Thus the problem of minimizing the sum of a smooth function H with a non-smooth one P
283 or R can be simplified to the following sub-problems ($l \in \{1, 2, \dots, N_c\}$):

284 (2.7)
$$\begin{cases} \mathbf{u}^{(k+1)} \in \underset{\mathbf{u}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{u} - (\mathbf{u}^{(k)} - \frac{1}{a} \nabla_{\mathbf{u}} H(\mathbf{u}^{(k)}, \mathbf{S}^{(k)}))\|_2^2 + \frac{\lambda}{a} P(\mathbf{u}), \\ \mathbf{s}_l^{(k+1)} \in \underset{\mathbf{s}_l}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{s}_l - (\mathbf{s}_l^{(k)} - \frac{1}{d} \nabla_{\mathbf{s}_l} H(\mathbf{u}^{(k+1)}, \mathbf{S}^{(k)}))\|_2^2 + \frac{\gamma}{d} R(\mathbf{s}_l). \end{cases}$$

285 **Definition 2.1 (Proximal operator [9]).** Given $\mathbf{x} \in \mathbb{C}^N$ and a lower semicontinuous function
286 $g : \mathbb{C}^N \rightarrow (-\infty, +\infty]$, the proximal operator associated to g is defined by:

287 (2.8)
$$\operatorname{prox}_{\tau g}(\mathbf{x}) := \underset{\mathbf{z}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 + \tau g(\mathbf{z}) : \mathbf{z} \in \mathbb{C}^N \right\}.$$

289 Using the definition of proximal operator in [Definition 2.1](#), we further obtain the following
290 Proximal Alternating Linearized Minimization (**PALM**) algorithm [9] at k -th iteration:

291 (2.9)
$$\begin{cases} \text{(U-Update): } \mathbf{u}^{(k+1)} \in \operatorname{prox}_{\frac{\lambda}{a} P}(\mathbf{u}^{(k)} - \frac{1}{a} \nabla_{\mathbf{u}} H(\mathbf{u}^{(k)}, \mathbf{S}^{(k)})), \\ \text{(S-Update): } \mathbf{s}_l^{(k+1)} \in \operatorname{prox}_{\frac{\gamma}{d} R}(\mathbf{s}_l^{(k)} - \frac{1}{d} \nabla_{\mathbf{s}_l} H(\mathbf{u}^{(k+1)}, \mathbf{S}^{(k)})), \end{cases}$$

292 where $l \in \{1, 2, \dots, N_c\}$ and $k \in \{0, 1, \dots, K-1\}$. $\mathbf{S}^{(0)} = \{\mathbf{s}_1^{(0)}, \mathbf{s}_2^{(0)}, \dots, \mathbf{s}_{N_c}^{(0)}\}$ is an initial
293 estimation of sensitivity maps. For each sensitivity map, e.g., $\mathbf{s}_l^{(0)} = \frac{\mathbf{A}^\top \mathbf{y}_l}{\sqrt{\sum_{l=1}^{N_c} (\mathbf{A}^\top \mathbf{y}_l)^* \odot (\mathbf{A}^\top \mathbf{y}_l)}}$,
294 the numerator is the initial estimation of the l -th coil MR image $\mathbf{A}^\top \mathbf{y}_l$ and the denominator
295 is the reconstructed MR image using the SoS. With $\mathbf{S}^{(0)}$, based on the PI imaging model,
296 $\mathbf{u}^{(0)} = \sum_{l=1}^{N_c} (\mathbf{s}_l^{(0)})^* \odot (\mathbf{A}^\top \mathbf{y}_l)$.

297 **Remark 2.2.** For the regularization functions in a specific form, their corresponding prox-
298 imal operators can be efficiently evaluated in closed-form formulations [41]. For more general
299 regularization functions, we may need to introduce an extra iterative optimization algorithm
300 to solve their corresponding proximal operators.

301 The above PALM algorithm consists of **Image Update (U-Update)** and **Sensitivity
302 Update (S-Update)** formulas. Using PALM for the CS-PI reconstruction model, it is chal-
303 lenging to choose appropriate regularization functions related to the priors of MR image and
304 coil sensitivities which often introduces visible artifacts in practice. It is also challenging to effi-
305 ciently compute the proximal operators corresponding to these regularizers, i.e., $\operatorname{prox}_{\frac{\lambda}{a} P}(\cdot)$ and

306 prox $\frac{\gamma}{2}R(\cdot)$, which usually involves another optimization algorithm. For example, it commonly
 307 takes hundreds of iterations to get a satisfactory reconstruction result for solving a reconstruc-
 308 tion model with two TV regularization terms. More importantly, it is also non-trivial to tune
 309 the algorithm hyper-parameters, such as a , d , α , β , λ , and γ in PALM, which will affect the
 310 convergence speed and accuracy of the algorithm. To overcome these difficulties, in this paper,
 311 we will design a deep learning method to learn the MR image and sensitivities regularizations
 312 and the hyper-parameters in PALM. We first design two learnable deep convolutional neural
 313 networks as sub-networks to replace the proximal operators in the algorithm, which implicitly
 314 learn the corresponding regularization functions of MR image and sensitivities. These two
 315 sub-networks directly learn the proximal mappings without relying on an iterative algorithm
 316 for solving proximal operators. Then, the PALM algorithm is unfolded as a deep architec-
 317 ture, whose parameters including the hyper-parameters of the iterations and the sub-networks
 318 can be learned by end-to-end training. If the objective function $E(\mathbf{u}, \mathbf{S})$ in Eqn. (2.1) is a
 319 Kurdyka–Łojasiewicz (KL) function, the whole generated sequence $\{(\mathbf{u}^{(k)}, \mathbf{S}^{(k)})\}$ in PALM
 320 algorithm can be proved to globally converge to a critical point $\{(\hat{\mathbf{u}}, \hat{\mathbf{S}})\}$ of the problem (**The-**
 321 **orem 1 in [9]**). In this work, we do not require two learnable regularization functions to
 322 satisfy the KL property. The convergence was proven by effectively integrating the Lipschitz
 323 properties of $H(\mathbf{u}, \mathbf{S})$ in the objective function and the Lipschitz properties of sub-networks,
 324 which can be potentially adapted to a wide class of non-convex problems.

325 2.3. Unrolled network for image and sensitivity priors learning. Since the regularization
 326 functions $P(\cdot)$ and $R(\cdot)$ in Eqn. (2.1) correspond to the proximal operators $\text{prox}_{\frac{\lambda}{\alpha}P}(\cdot)$ and
 327 $\text{prox}_{\frac{\gamma}{d}R}(\cdot)$ in Eqn. (2.9), we substitute these proximal operators by two designed deep con-
 328 volutional neural networks (CNNs) $F_u(\cdot; \theta_u)$, $F_s(\cdot; \theta_s) : \mathbb{C}^N \rightarrow \mathbb{C}^N$, named **U-Subnet** and
 329 **S-Subnet** in each iteration of PALM, i.e.,

$$330 \quad (2.10) \quad (\textbf{U-Update}) : \quad \mathbf{u}^{(k+1)} = F_u(\tilde{\mathbf{u}}^{(k)}; \theta_u),$$

$$(2.11) \quad (\text{S-Update}) : \quad \mathbf{s}_l^{(k+1)} = F_s(\tilde{\mathbf{s}}_l^{(k)}; \theta_s),$$

333 where $l \in \{1, 2, \dots, N_c\}$, θ_u and θ_s are parameters of sub-nets. Let $\eta_u = \frac{1}{a}$ and $\eta_s = \frac{1}{d}$, then

$$334 \quad (2.12) \quad \tilde{\mathbf{u}}^{(k)} = \mathbf{u}^{(k)} - \eta_u \nabla_{\mathbf{u}} H(\mathbf{u}^{(k)}, \mathbf{S}^{(k)}), \quad \tilde{\mathbf{s}}_l^{(k)} = \mathbf{s}_l^{(k)} - \eta_s \nabla_{\mathbf{s}_l} H(\mathbf{u}^{(k+1)}, \mathbf{S}^{(k)}).$$

335 An iterative PALM naturally defines a non-linear reconstruction mapping to estimate the
 336 reconstructed image and sensitivity maps from measurements of multi-receivers. In order
 337 to learn sub-networks and hyper-parameters in the algorithm instead of setting them by
 338 hand, we unroll the above PALM iterations to be a deep architecture, dubbed as BPMRI-
 339 Net. As shown in Figure 1(a), the proposed BPMRI-Net contains multiple iterations, each
 340 of which corresponds to one iteration of the PALM algorithm. Given a set of multi-channel
 341 measurements in k -space (i.e., $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_c}\}$), it flows over the network iteration
 342 by iteration to correct the estimated MR images (i.e., $\mathbf{u}^{(k)}$) and sensitivity maps (i.e., $\mathbf{S}^{(k)}$),
 343 and automatically generates the optimal reconstructed MR image (i.e., $\hat{\mathbf{u}}$). As illustrated
 344 in Figure 1(b), each iteration of the network consists of **Image Update block (U-Update)**

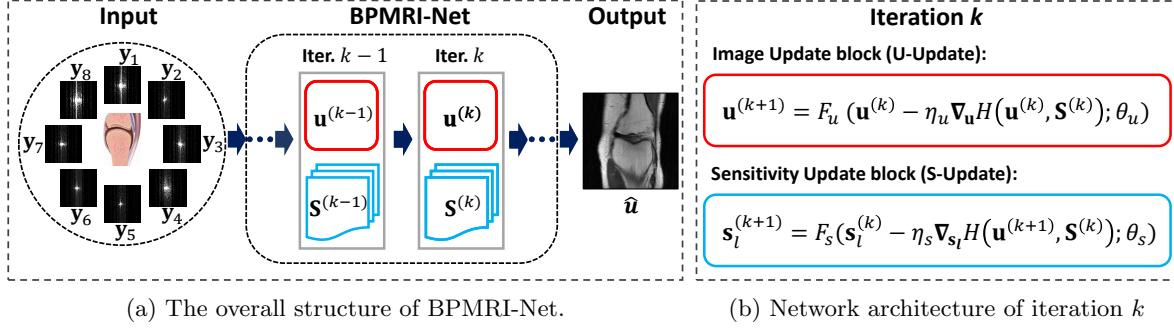


Figure 1: (a) Given an input set of multi-channel measurements in k -space, it flows over the network iteration by iteration to update the MR images and sensitivity maps, then the BPMRI-Net outputs the optimal MR image. (b) Each iteration of the BPMRI-Net consists of an image update block (**U-Update**) and a sensitivity update block (**S-Update**).

345 and **Sensitivity Update block (S-Update)**, which implement Eqn. (2.10) and Eqn. (2.11)
 346 for updating the reconstructed MR image and sensitivity maps, respectively. Each update
 347 block contains a partial gradient descent step with respect to the function H and a sub-
 348 network. Then, parameters from two sub-networks and hyper-parameters in iterations are
 349 transformed into parameters of the proposed BPMRI-Net which can be learned end-to-end
 350 because the whole process of iteration is differentiable.

351 **The structure of two sub-networks.** As shown in Figure 2, **U-Subnet** is designed as a
 352 deep CNN including 9 cascaded convolutional (Conv) blocks and a Conv layer with extra skip
 353 connection for residual learning, which outputs a reconstructed complex-valued MR image.
 354 Each Conv block consists of a Conv layer, an Instance Normalization (IN) layer, and a ReLU
 355 activation layer. **S-Subnet** has a similar structure with two Conv blocks and an extra Conv
 356 layer which outputs the estimated sensitivity map. In **U-Subnet**, each Conv layer in Conv
 357 blocks has 64 filters in size of 3×3 . The final Conv layer has 2 filters in size of 3×3 . In
 358 **S-Subnet**, each Conv layer in Conv blocks has 32 filters in size of 5×5 . The extra Conv
 359 layer has 2 filters in size of 9×9 . To keep boundedness, the outputs of two subnetworks are
 360 bounded by additional projection Π_{B_r} with radius r , i.e.,

$$361 \quad (2.13) \quad \Pi_{B_r}(\mathbf{x}) = \begin{cases} r \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \|\mathbf{x}\| \geq r, \\ \mathbf{x}, & \|\mathbf{x}\| < r. \end{cases}$$

362 Furthermore, in order to make the proposed sub-networks satisfy the Lipschitz continuity for
 363 stability and convergence, we perform the soft spectral normalization on the convolutional
 364 operator in each layer of the sub-networks. The output feature map of i -th channel of a
 365 convolutional (Conv) layer is given by:

$$366 \quad (2.14) \quad \phi_i(\mathbf{x}) = \sum_{m=1}^{N_f} \mathbf{W}_{i,m} \mathbf{x}_m + \mathbf{b}_i,$$

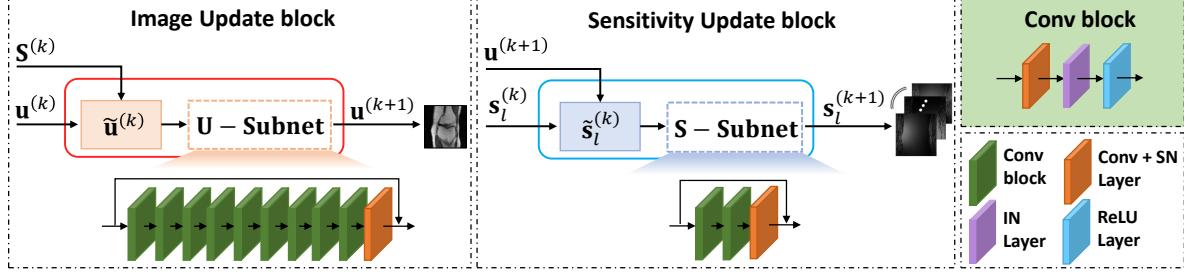


Figure 2: The structure of image update block and sensitivity update block.

367 where \mathbf{x}_m is the input feature map of m -th channel and $\mathbf{W}_{i,m}$ is a matrix corresponding to a
 368 convolution filter $\omega_{i,m}$. N_f is the number of feature maps produced by the previous layer and
 369 b_i is the i -th bias. Thus, we normalize the weight matrix $\mathbf{W}_{i,m}$ via

370 (2.15)
$$\tilde{\mathbf{W}}_{i,m} = \begin{cases} \frac{c\mathbf{W}_{i,m}}{\|\mathbf{W}_{i,m}\|}, & \text{if } \frac{c}{\|\mathbf{W}_{i,m}\|} < 1; \\ \mathbf{W}_{i,m}, & \text{else,} \end{cases}$$

371 where $c < 1$ is a scaling coefficient. $\|\mathbf{W}_{i,m}\|$ is the spectral norm (maximum singular value)
 372 of $\mathbf{W}_{i,m}$ which can be estimated by performing power-iteration [17].

373 **Network training.** We use the following loss function to train the BPMRI-Net:

374 (2.16)
$$\ell(\theta) = \sum_n \|\hat{\mathbf{u}}_n(\theta) - \mathbf{u}_n^{gt}\|_2^2,$$

375 where $\hat{\mathbf{u}}_n(\theta)$ is the estimated MR image by the network for n -th training sample. The \mathbf{u}_n^{gt} is
 376 the target MR image reconstructed from the fully-sampled k -space data, which is synthesized
 377 based on the sensitivity maps estimated by ESPIRiT using sufficient samples of the k -space
 378 center (e.g., 80 ACS lines). The loss function does not include supervisory loss of sensitivities.
 379 θ denotes the parameters of BPMRI-Net including the parameters of two sub-networks, i.e.,
 380 θ_u and θ_s , and the hyper-parameters η_u , η_s , α , and β , which are shared in different iterations.
 381 The gradients of loss w.r.t. the network parameters can be computed by auto-differentiation
 382 and we learn these parameters using Adam with a learning rate of 1e-4 by PyTorch. In order to
 383 constrain the learned hyper-parameters to be non-negative, we set $\eta_u = \exp(\eta'_u)$, $\eta_s = \exp(\eta'_s)$,
 384 $\alpha = \exp(\alpha')$, and $\beta = \exp(\beta')$. Therefore, in the process of training, the parameters η_u , η_s ,
 385 α , and β are updated indirectly by updating the parameters η'_u , η'_s , α' , and β' . We train the
 386 network for 120 epochs. After training, given multi-coil under-sampled k -space data, we can
 387 reconstruct the final MR image by [Algorithm 2.1](#).

388 **Remark 2.3.** Instead of directly learning the regularization functions, we implicitly learn
 389 its proximal operators in the iterative algorithm with deep networks. We cannot theoretically
 390 guarantee the existence of the regularization terms corresponding to the learned sub-networks.
 391 Some existing methods learn regularization explicitly, however, their networks are assumed to
 392 obey some specific properties [44] such as Jacobian Symmetry, Strong Passivity, etc., which

Algorithm 2.1 Testing phase of BPMRI-Net

Input : Iteration number K ; initialization $\mathbf{S}^{(0)} = \{\mathbf{s}_1^{(0)}, \mathbf{s}_2^{(0)}, \dots, \mathbf{s}_{N_c}^{(0)}\}$ and $\mathbf{u}^{(0)} = \sum_{l=1}^{N_c} (\mathbf{s}_l^{(0)})^* \odot (\mathbf{A}^H \mathbf{y}_l)$; trained network parameters $\theta_u, \theta_s, \eta_u, \eta_s, \alpha$ and β .

for k from 0 to $K - 1$ **do**

$$\mathbf{u}^{(k+1)} = F_u \left(\mathbf{u}^{(k)} - \eta_u \nabla_{\mathbf{u}} H(\mathbf{u}^{(k)}, \mathbf{S}^{(k)}; \theta_u) \right),$$

$$\mathbf{s}_l^{(k+1)} = F_s \left(\mathbf{s}_l^{(k)} - \eta_s \nabla_{\mathbf{s}_l} H(\mathbf{u}^{(k+1)}, \mathbf{S}^{(k)}; \theta_s) \right), \quad l \in \{1, \dots, N_c\}.$$

end for

Output: $\mathbf{u}^{(K)}$.

393 greatly constrain the representation ability of the network. Thus, most of the methods consider
 394 regularization learning of particular forms, such as linear denoiser [38] and regularization
 395 term in the form of summation [13, 47, 66]. Recently, the regularization by denoising (RED)
 396 approaches [2, 44] and the minimum mean square error (MMSE) denoiser [62] have been
 397 proposed, which can only provide approximations of the gradient of regularizer and MMSE
 398 regressors respectively.

399 **3. Convergence.** For the convergence analysis of the proposed algorithm, we first intro-
 400 duce a definition, two mild assumptions, and two conditions on this problem. For simplicity
 401 of notation, we define $H_l(\mathbf{u}, \mathbf{s}_l) \triangleq \frac{1}{2} \|\mathbf{A}(\mathbf{s}_l \odot \mathbf{u}) - \mathbf{y}_l\|^2 + \frac{\beta}{2} \|\mathbf{s}_l\|^2$ ($l \in \{1, 2, \dots, N_c\}$), and $\|\cdot\|$
 402 represents a ℓ_2 -norm of a vector or a spectral norm of a matrix.

403 **Definition 3.1 (Split Norm).** For all $k \in \mathbb{N}$, we define

$$404 \quad \mathbf{z}^{(k)} \triangleq ((\mathbf{u}^{(k)})^\top, (\mathbf{s}_1^{(k)})^\top, (\mathbf{s}_2^{(k)})^\top, \dots, (\mathbf{s}_{N_c}^{(k)})^\top)^\top \in \mathbb{C}^{(N_c+1)N}$$

406 and we define its “split norm” to be $\|\mathbf{z}^{(k)}\|_{\text{split}} = \|\mathbf{u}^{(k)}\| + \sum_{l=1}^{N_c} \|\mathbf{s}_l^{(k)}\|$.

407 **Property 3.2.** The defined “split norm” is a norm.

408 **Proof.** For all \mathbf{z} and $\mathbf{z}' \in \mathbb{C}^{(N_c+1)N}$, we prove the proprieties of subadditivity, absolute
 409 homogeneity, positive definiteness, and non-negativity for the split norm as follows. (1) Sub-
 410 additivity:

$$411 \quad \begin{aligned} \|\mathbf{z} + \mathbf{z}'\|_{\text{split}} &= \|\mathbf{u} + \mathbf{u}'\| + \sum_{l=1}^{N_c} \|\mathbf{s}_l + \mathbf{s}'_l\| \\ 412 &\leq \|\mathbf{u}\| + \|\mathbf{u}'\| + \sum_{l=1}^{N_c} (\|\mathbf{s}_l\| + \|\mathbf{s}'_l\|) \\ 413 &= \|\mathbf{z}\|_{\text{split}} + \|\mathbf{z}'\|_{\text{split}}. \end{aligned}$$

415 (2) Absolute homogeneity: $\|s\mathbf{z}\|_{\text{split}} = \|s\mathbf{u}\| + \sum_{l=1}^{N_c} \|s\mathbf{s}_l\| = |s| \|\mathbf{z}\|_{\text{split}}$, for all scalars s . (3)
 416 Positive definiteness: If $\|\mathbf{z}\|_{\text{split}} = \|\mathbf{u}\| + \sum_{l=1}^{N_c} \|\mathbf{s}_l\| = 0$, we have $\mathbf{u} = \mathbf{0}$ and $\mathbf{s}_l = \mathbf{0}$, hence
 417 $\mathbf{z} = \mathbf{0}$. (4) Non-negativity: $\|\mathbf{z}\|_{\text{split}} = \|\mathbf{u}\| + \sum_{l=1}^{N_c} \|\mathbf{s}_l\| \geq 0$. Thus, the “split norm” is a norm. ■

418 Property 3.3. For any vector $\mathbf{z} \in \mathbb{R}^{(N_c+1)N'}$, we have $\|\mathbf{z}\| \leq \|\mathbf{z}\|_{\text{split}}$. This conclusion can
419 be proved based on $\|\mathbf{u}\|^2 + \sum_{l=1} \|\mathbf{s}_l\|^2 \leq (\|\mathbf{u}\| + \sum_{l=1} \|\mathbf{s}_l\|)^2$.

420 Assumption 3.4 (Bounded Norm). For all $k \in \mathbb{N}$, the norms of $\mathbf{u}^{(k)}$ and $\mathbf{s}_l^{(k)}$ are bounded
421 ($l \in \{1, 2, \dots, N_c\}$), i.e., there exist positive number B_u and B_s such that

$$422 \quad \|\mathbf{u}^{(k)}\| \leq B_u, \quad \|\mathbf{s}_l^{(k)}\| \leq B_s.$$

424 Remark 3.5. To satisfy the assumption of “Bounded Norm”, we attach a bounded projec-
425 tion transform Π_{B_r} in Eqn. (2.13) to the outputs of both sub-networks, respectively.

426 Assumption 3.6 (Network Lipschitz Continuous). There exists a constant number $0 < L < 1$
427 such that F_s and F_u are $(1 + L)$ -Lipschitzian.

428 Remark 3.7. Since we perform soft spectrum normalization on each convolutional operator
429 of the two sub-networks using Eqn. (2.15), and their network structures include skip connec-
430 tions for residual learning, two designed sub-networks can be represented as $F_u = Id + \tilde{F}_u$
431 and $F_s = Id + \tilde{F}_s$, where \tilde{F}_u and \tilde{F}_s are L -Lipschitzian ($0 < L < 1$). Thus, F_s and F_u are
432 $(1 + L)$ -Lipschitzian satisfying the above assumption.

433 Condition 3.8. Parameters α and β satisfy

$$434 \quad (3.1) \quad \alpha > \max \left\{ \frac{N_c B_s^2 \|\mathbf{A}\|^2 (2 + 3L)}{2}, (2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \|\mathbf{A}^\top \mathbf{y}_l\|)(2 + 3L) \right\},$$

$$435 \quad (3.2) \quad \beta > \max \left\{ \frac{B_u^2 \|\mathbf{A}\|^2 (2 + 3L)}{2}, (2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \|\mathbf{A}^\top \mathbf{y}_l\|)(2 + 3L) N_c \right\}.$$

437 Remark 3.9. When training network parameters α and β , we update the parameters to
438 make them larger than the pre-estimated lower bounds plus a small positive number. These
439 two lower bounds increase with the increase of Lipschitz constant L .

440 Condition 3.10. Parameters η_u and η_s satisfy

$$441 \quad (3.3) \quad \frac{2 + 3L}{3\alpha(1 + L)} < \eta_u < \min \left\{ \frac{4 + 3L}{3L_{uu}(1 + L)}, \frac{1}{3(1 + L)L_{su}} \right\},$$

$$442 \quad (3.4) \quad \frac{2 + 3L}{3\beta(1 + L)} < \eta_s < \min \left\{ \frac{4 + 3L}{3L_{ss}(1 + L)}, \frac{1}{3(1 + L)L_{su}N_c} \right\},$$

444 where $L_{su} = 2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \|\mathbf{A}^\top \mathbf{y}_l\|$, $L_{ss} = B_u^2 \|\mathbf{A}\|^2 + \beta$, $L_{uu} = N_c B_s^2 \|\mathbf{A}\|^2 + \alpha$.

445 Remark 3.11. When training the network parameters η_u and η_s , we project the parameters
446 into the intervals defined in Eqns. (3.3) and (3.4) to satisfy the constraints.

447 Based on the above assumptions and conditions, we can prove that the proposed **Blind-**
448 **PMRI-Net**, as an iterative algorithm, can converge to a fixed point asymptotically ¹.

¹Note that the concept of convergence discussed in this work is about the output of each iteration of the network, and is not related to the training of network parameters.

449 There are three main challenges to analyzing the convergence of BPMRI-Net. **(1)** Since
 450 the proximal operators are replaced by deep convolutional networks, we can not exploit the
 451 first-order optimality condition of the optimization problem in Eqn. (2.8) in the proofs, which
 452 was often used in convergence analysis of proximal optimization related algorithms [5, 6, 9].
 453 **(2)** Because both sub-networks based on residual learning are $(1+L)$ -Lipschitzian ($0 < L < 1$),
 454 they are not non-expansive operators. Therefore, the fixed point theorem cannot be directly
 455 used to ensure convergence. **(3)** Iterations in Eqns. (2.10) and (2.11) involve optimization
 456 of two variables \mathbf{u} and \mathbf{S} in an alternative way. This brings a challenge that the update of
 457 one variable will directly affect the update of the other one. Our key idea is that, under the
 458 assumptions mentioned above, we are able to derive the Lipschitz properties of the gradients
 459 of function H w.r.t. \mathbf{u} and \mathbf{s}_l (see Lemma 3.12). Then combining these Lipschitz properties
 460 with Lemma 3.14, we are able to obtain that sequence $\{\mathbf{z}_k\}$ is **contractive** in “split norm”
 461 (see Lemma 3.15), which enables us to use Property 3.3 and Banach contraction principle to
 462 prove the convergence in ℓ_2 -norm (see Theorem 3.16).

463 **Lemma 3.12.** *Under Assumption 3.4, $H(\mathbf{u}, \mathbf{S})$ in Eqn. (2.2) is strongly-convex with respect
 464 to \mathbf{u} with constant α and strongly-convex with respect to \mathbf{s}_l with constant β , and we have*

465 (3.5)
$$\|\nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}_1) - \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}_2)\| \leq L_{ss} \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\|,$$

466 (3.6)
$$\|\nabla_{\mathbf{u}} H(\mathbf{u}_1, \mathbf{S}) - \nabla_{\mathbf{u}} H(\mathbf{u}_2, \mathbf{S})\| \leq L_{uu} \|\mathbf{u}_1 - \mathbf{u}_2\|,$$

467 (3.7)
$$\|\nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}_1) - \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}_2)\| \leq L_{su} \sum_{l=1}^{N_c} \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\|,$$

468 (3.8)
$$\|\nabla_{\mathbf{s}_l} H(\mathbf{u}_1, \mathbf{S}) - \nabla_{\mathbf{s}_l} H(\mathbf{u}_2, \mathbf{S})\| \leq L_{su} \|\mathbf{u}_1 - \mathbf{u}_2\|,$$

470 for all $\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{s}_l, \mathbf{s}_{l,1}, \mathbf{s}_{l,2} \in \mathbb{C}^N, l \in \{1, 2, \dots, N_c\}$, and $\mathbf{S}, \mathbf{S}_1, \mathbf{S}_2 \in \mathbb{C}^{N \times N_c}$, where $\alpha > 0$,
 471 $\beta > 0$, $\mathbf{S}_1 = \{\mathbf{s}_{l,1}\}_l$, $\mathbf{S}_2 = \{\mathbf{s}_{l,2}\}_l$, $L_{ss} = B_u^2 \|\mathbf{A}\|^2 + \beta$, $L_{uu} = N_c B_s^2 \|\mathbf{A}\|^2 + \alpha$ and $L_{su} =$
 472 $2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \|\mathbf{A}^\top \mathbf{y}_l\|$.

473 *Proof.* See Appendix A. ■

474 **Remark 3.13.** These Lipschitz coefficients depend on the boundedness of the norms of $\mathbf{u}^{(k)}$
 475 and $\mathbf{s}_l^{(k)}$ in each iteration, which are controlled by the projection Π_{B_r} .

476 **Lemma 3.14 (Corollary of Theorem 2.1 in [53] and Lemma 7 in [45]).** *Assume f is strongly
 477 convex with constant μ ($\mu > 0$), and ∇f is Lipschitz continuous with constant ρ ($\rho > 0$), then
 478 for any $\mathbf{x}, \mathbf{z} \in \mathbb{C}^N$ and $\zeta > 0$, we have*

479
$$\|(I - \zeta \nabla f)(\mathbf{x}) - (I - \zeta \nabla f)(\mathbf{z})\| \leq \max\{|1 - \zeta \mu|, |1 - \zeta \rho|\} \|\mathbf{x} - \mathbf{z}\|.$$

481 *Proof.* Because f is strongly convex with constant μ and ∇f is Lipschitz continuous with
 482 constant ρ , we have

483
$$\mu \mathbf{I} \leq \mathbf{H}_f \leq \rho \mathbf{I},$$

485 where \mathbf{H}_f denotes the Hessian of f . In other words, the maximum singular value of \mathbf{H}_f is less
 486 than or equal to ρ and the minimum singular value is larger than or equal to μ . Therefore,

487 the singular values of $I - \zeta \mathbf{H}_f$ are between values $1 - \zeta\rho$ and $1 - \zeta\mu$. Hence we have

$$488 \quad \|I - \zeta \mathbf{H}_f\|_2 \leq \max\{|1 - \zeta\mu|, |1 - \zeta\rho|\}.$$

490 Then we can get the conclusion in [Lemma 3.14](#). ■

491 **Lemma 3.15.** *Under [Assumptions 3.4](#) and [3.6](#) and [Conditions 3.8](#) and [3.10](#), there exists*
 492 *$0 < q < 1$ such that for all $k \in \mathbb{N}$,*

$$493 \quad 494 \quad \left\| \mathbf{z}^{(k+2)} - \mathbf{z}^{(k+1)} \right\|_{split} \leq q \left\| \mathbf{z}^{(k+1)} - \mathbf{z}^{(k)} \right\|_{split}.$$

495 **Proof.** For all $k \in \mathbb{N}$, $\mathbf{z}^{(k)} = ((\mathbf{u}^{(k)})^\top, (\mathbf{s}_1^{(k)})^\top, (\mathbf{s}_2^{(k)})^\top, \dots, (\mathbf{s}_{N_c}^{(k)})^\top)^\top$. Firstly, considering $\mathbf{s}_l^{(k+2)}$ and $\mathbf{s}_l^{(k+1)}$ in Eqn. [\(2.11\)](#), it is straightforward to check that $\nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}) = \nabla_{\mathbf{s}_l} H_l(\mathbf{u}, \mathbf{s}_l)$. Hence we have

$$498 \quad \mathbf{s}_l^{(k+2)} = F_s(I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+2)}, \cdot))(\mathbf{s}_l^{(k+1)}), \\ 499 \quad \mathbf{s}_l^{(k+1)} = F_s(I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+1)}, \cdot))(\mathbf{s}_l^{(k)}).$$

501 Therefore, by [Assumption 3.6](#), we have

$$502 \quad \left\| \mathbf{s}_l^{(k+2)} - \mathbf{s}_l^{(k+1)} \right\| \\ 503 \quad \leq (1 + L) \left\| (I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+2)}, \cdot))(\mathbf{s}_l^{(k+1)}) - (I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+1)}, \cdot))(\mathbf{s}_l^{(k)}) \right\| \\ 504 \quad \leq (1 + L) \left\| (I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+2)}, \cdot))(\mathbf{s}_l^{(k+1)}) - (I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+2)}, \cdot))(\mathbf{s}_l^{(k)}) \right\| \\ 505 \quad + (1 + L) \left\| (I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+2)}, \cdot))(\mathbf{s}_l^{(k)}) - (I - \eta_s \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+1)}, \cdot))(\mathbf{s}_l^{(k)}) \right\| \\ 506 \quad \stackrel{(i)}{\leq} (1 + L) \max\{|1 - \eta_s\beta|, |1 - \eta_s L_{ss}|\} \left\| \mathbf{s}_l^{(k+1)} - \mathbf{s}_l^{(k)} \right\| \\ 507 \quad + \eta_s (1 + L) \left\| \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+2)}, \mathbf{s}_l^{(k)}) - \nabla_{\mathbf{s}_l} H_l(\mathbf{u}^{(k+1)}, \mathbf{s}_l^{(k)}) \right\| \\ 508 \quad \stackrel{(ii)}{\leq} (1 + L) \left[\max\{|1 - \eta_s\beta|, |1 - \eta_s L_{ss}|\} \left\| \mathbf{s}_l^{(k+1)} - \mathbf{s}_l^{(k)} \right\| + \eta_s L_{su} \left\| \mathbf{u}^{(k+2)} - \mathbf{u}^{(k+1)} \right\| \right],$$

510 where (i) combines [Lemmas 3.12](#) and [3.14](#), and (ii) comes from [Lemma 3.12](#).

511 Secondly, considering $\mathbf{u}^{(k+2)}$ and $\mathbf{u}^{(k+1)}$, we obtain

$$512 \quad \mathbf{u}^{(k+2)} = F_u(I - \eta_u \nabla_{\mathbf{u}} H(\cdot, \mathbf{S}^{(k+1)}))(\mathbf{u}^{(k+1)}), \\ 513 \quad \mathbf{u}^{(k+1)} = F_u(I - \eta_u \nabla_{\mathbf{u}} H(\cdot, \mathbf{S}^{(k)}))(\mathbf{u}^{(k)}).$$

515 Similar to Eqn. [\(3.9\)](#), we get

$$516 \quad \left\| \mathbf{u}^{(k+2)} - \mathbf{u}^{(k+1)} \right\| \\ 517 \quad \leq (1 + L) \left[\max\{|1 - \eta_u\alpha|, |1 - \eta_u L_{uu}|\} \left\| \mathbf{u}^{(k+1)} - \mathbf{u}^{(k)} \right\| + \eta_u L_{su} \sum_{l=1}^{N_c} \left\| \mathbf{s}_l^{(k+1)} - \mathbf{s}_l^{(k)} \right\| \right].$$

519 Thirdly, combining Eqns. (3.9) and (3.10), we have

$$\begin{aligned} 520 \quad & \sum_{l=1}^{N_c} \left\| \mathbf{s}_l^{(k+2)} - \mathbf{s}_l^{(k+1)} \right\| + (1 - \eta_s L_{su}(1 + L)N_c) \left\| \mathbf{u}^{(k+2)} - \mathbf{u}^{(k+1)} \right\| \\ 521 \quad & \leq (1 + L)(\eta_u L_{su} + \max\{|1 - \eta_s \beta|, |1 - \eta_s L_{ss}|)\} \sum_{l=1}^{N_c} \left\| \mathbf{s}_l^{(k+1)} - \mathbf{s}_l^{(k)} \right\| \\ 522 \quad & + (1 + L) \max\{|1 - \eta_u \alpha|, |1 - \eta_u L_{uu}|\} \left\| \mathbf{u}^{(k+1)} - \mathbf{u}^{(k)} \right\|. \end{aligned}$$

524 When $\eta_s < \frac{1}{3(1+L)L_{su}N_c}$, we obtain $1 - \eta_s L_{su}(1 + L)N_c > 0$ and

$$\begin{aligned} 525 \quad & \sum_{l=1}^{N_c} \left\| \mathbf{s}_l^{(k+2)} - \mathbf{s}_l^{(k+1)} \right\| + \left\| \mathbf{u}^{(k+2)} - \mathbf{u}^{(k+1)} \right\| \leq q \left(\sum_{l=1}^{N_c} \left\| \mathbf{s}_l^{(k+1)} - \mathbf{s}_l^{(k)} \right\| + \left\| \mathbf{u}^{(k+1)} - \mathbf{u}^{(k)} \right\| \right), \\ 526 \end{aligned}$$

527 where $q = \frac{(1+L)(\eta_u L_{su} + \max\{|1 - \eta_u \alpha|, |1 - \eta_u L_{uu}|, |1 - \eta_s \beta|, |1 - \eta_s L_{ss}|\})}{1 - \eta_s L_{su}(1 + L)N_c}$. When the following condition

$$528 \quad (3.11) \quad \eta_s L_{su}N_c + \eta_u L_{su} + \max\{|1 - \eta_u \alpha|, |1 - \eta_u L_{uu}|, |1 - \eta_s \beta|, |1 - \eta_s L_{ss}|\} < \frac{1}{1 + L}$$

530 is satisfied, the contraction coefficient $q < 1$.

531 Finally, to make the above condition hold, let the three items on the left-hand side of
532 Eqn. (3.11) to be less than $\frac{1}{3(1+L)}$, and we get

$$533 \quad \frac{2 + 3L}{3\alpha(1 + L)} < \eta_u < \min \left\{ \frac{4 + 3L}{3L_{uu}(1 + L)}, \frac{1}{3(1 + L)L_{su}} \right\},$$

$$534 \quad \frac{2 + 3L}{3\beta(1 + L)} < \eta_s < \min \left\{ \frac{4 + 3L}{3L_{ss}(1 + L)}, \frac{1}{3(1 + L)L_{su}N_c} \right\}.$$

536 To make sure that the above inequalities are feasible, we further need to ensure $\frac{2+3L}{3\alpha(1+L)} <$
537 $\min\{\frac{4+3L}{3L_{uu}(1+L)}, \frac{1}{3(1+L)L_{su}}\}$ and $\frac{2+3L}{3\beta(1+L)} < \min\{\frac{4+3L}{3L_{ss}(1+L)}, \frac{1}{3(1+L)L_{su}N_c}\}$. Then, we get the re-
538 strictions of α and β :

$$539 \quad \alpha > \max \left\{ \frac{N_c B_s^2 \|\mathbf{A}\|^2 (2 + 3L)}{2}, (2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \|\mathbf{A}^\top \mathbf{y}_l\|) (2 + 3L) \right\},$$

$$540 \quad \beta > \max \left\{ \frac{B_u^2 \|\mathbf{A}\|^2 (2 + 3L)}{2}, (2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \|\mathbf{A}^\top \mathbf{y}_l\|) (2 + 3L) N_c \right\}. \quad \blacksquare$$

542 Based on the above assumptions, conditions, and Lemmas, our main result (i.e., The-
543 rem 3.16) is that **BPMRI-Net** as an iterative algorithm, can linearly converge to a fixed
544 point asymptotically. We also experimentally verify the convergence property of BPMRI-Net
545 in subsection 4.4. Note that the following theorem proves the convergence of the sequence gen-
546 erated by BPMRI-Net, but we can not guarantee that the sequences generated with different
547 initialization converge to the same fixed point.

548 **Theorem 3.16.** Under Assumptions 3.4 and 3.6 and Conditions 3.8 and 3.10, there exists
 549 a point $((\hat{\mathbf{u}})^\top, (\hat{\mathbf{s}}_1)^\top, (\hat{\mathbf{s}}_2)^\top, \dots, (\hat{\mathbf{s}}_{N_c})^\top)^\top$, such that the sequence from BPMRI-Net

550 $((\mathbf{u}^{(k)})^\top, (\mathbf{s}_1^{(k)})^\top, (\mathbf{s}_2^{(k)})^\top, \dots, (\mathbf{s}_{N_c}^{(k)})^\top)^\top \rightarrow ((\hat{\mathbf{u}})^\top, (\hat{\mathbf{s}}_1)^\top, (\hat{\mathbf{s}}_2)^\top, \dots, (\hat{\mathbf{s}}_{N_c})^\top)^\top$.

551 The convergence error exponentially decays with iterations:

552
$$\left\| ((\mathbf{u}^{(k)})^\top, (\mathbf{s}_1^{(k)})^\top, (\mathbf{s}_2^{(k)})^\top, \dots, (\mathbf{s}_{N_c}^{(k)})^\top)^\top - ((\hat{\mathbf{u}})^\top, (\hat{\mathbf{s}}_1)^\top, (\hat{\mathbf{s}}_2)^\top, \dots, (\hat{\mathbf{s}}_{N_c})^\top)^\top \right\|$$

 553 (3.12) $\leq \frac{q^k}{1-q} \left(\|\mathbf{u}^{(1)} - \mathbf{u}^{(0)}\| + \sum_{l=1}^{N_c} \|\mathbf{s}_l^{(1)} - \mathbf{s}_l^{(0)}\| \right),$

555 where $q = \frac{(1+L)(\eta_u L_{su} + \max\{|1-\eta_u \alpha|, |1-\eta_u L_{uu}|, |1-\eta_s \beta|, |1-\eta_s L_{ss}|\})}{1-\eta_s L_{su}(1+L)N_c}$ and $0 < q < 1$.

556 *Proof.* See Appendix B. ■

557 **4. Experiments.** In this section, we first evaluate the performance of different methods for
 558 parallel imaging on a public knee dataset. We will show that the proposed BPMRI-Net out-
 559 performs the reconstruction accuracy of the current PI methods. Secondly, we experimentally
 560 evaluate the effect of the number of ACS lines and the convergence property of BPMRI-Net.
 561 Thirdly, we empirically analyze the influence of image update block and sensitivity update
 562 block on network performance by ablation study. Finally, we report reconstruction results on
 563 our collected brain datasets.

564 **4.1. Experimental settings.** We train and test BPMRI-Net on in-vivo knee MR data
 565 from NYU dataset ² and two in-vivo brain datasets that we collected. The knee dataset
 566 used a standard clinical protocol to obtain data for differences in anatomy, gender, pathology,
 567 age, and body mass index, which consists of three 2D turbo spin echo (TSE) sequences with
 568 different contrasts including proton-density (PD), fat-saturated PD, and fat-saturated T2.
 569 The raw data consists of 15-channel k -space data with readout-direction oversampling. The
 570 images are cropped to the central 320×320 pixel region for visualization. We randomly
 571 selected 10 slices from each of the 55 subjects and finally obtained 350, 100, and 100 multi-
 572 channel data as training, validation, and test sets. For the brain datasets, we collect T1-
 573 weighted raw data by a 3T Siemens Spectra scanner using a 12-channel receive head coil
 574 and a 8-channel receive head coil from 25 subjects, respectively. For each brain dataset, we
 575 randomly select 20 subjects for training, 5 subjects for validation, and 5 subjects for testing,
 576 respectively providing 300, 100, and 100 multi-channel data. The scan protocol is the 3D
 577 MPRAGE sequence with TR=1880 ms, TE=3.9 ms, flip angle=8°, volume size=256×256×128
 578 and voxel dimension=1 mm×1 mm×1 mm. All data were collected without acceleration
 579 and oversampling. In experiments, we use two types of 1D Cartesian under-sampling masks
 580 including uniform masks and random masks, which are popular and practical CS-PI masks as
 581 shown in Figure 3 with acceleration factors of $R_A = 4, 6$, and 8, respectively.

582 Some auto-calibration and calibration-less methods, such as GRAPPA and OT-cycleGAN,
 583 reconstruct multi-channel sensitivity encoded images and combine them using the SoS oper-
 584 ation to generate a single-channel real-valued MR image. For fair comparisons with the

²<http://mridata.org>

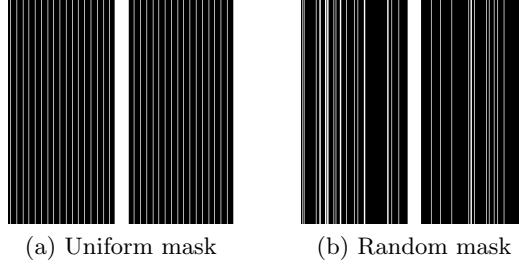


Figure 3: Two examples of 1D Cartesian under-sampling masks at $8 \times$ acceleration in k -space.

585 methods of reconstructing real-valued MR images, we generate real-valued ground-truth MR
 586 images by combining multi-channel images reconstructed from the fully-sampled k -space data
 587 using the SoS operation, i.e., $\mathbf{u}^{gt} = \sqrt{\sum_{l=1}^{N_c} \mathbf{v}_l^* \odot \mathbf{v}_l}$. The quality of the reconstructed MR
 588 images is measured by the average peak signal-to-noise ratio (PSNR), normalized root-mean-
 589 square error (nRMSE), and structural similarity index measure (SSIM) over the entire test
 590 set. These quantitative metrics are calculated based on the magnitude of the ground-truth
 591 MR image \mathbf{u}^{gt} and the reconstructed MR image $\hat{\mathbf{u}}$. For example, nRMSE = $\frac{\|\mathbf{u}^{gt} - \hat{\mathbf{u}}\|_2}{\|\mathbf{u}^{gt}\|_2}$, where
 592 $|\cdot|$ is the element-wise magnitude of each element in a complex-valued vector. In addition,
 593 other methods such as Learning VN and MoDL, output complex-valued MR images based on
 594 sensitivity reconstruction. To compare with these methods, we generate additional complex-
 595 valued ground-truth MR images by combining multi-channel images reconstructed from the
 596 fully-sampled k -space data based on the sensitivity maps estimated by ESPIRiT using 80
 597 ACS lines, i.e., $\mathbf{u}^{gt} = \sqrt{\sum_{l=1}^{N_c} \mathbf{s}_l^* \odot \mathbf{v}_l}$. We calculate the nRMSE value of the complex-valued
 598 ground-truth MR image and the reconstructed MR image, denoted as $nRMSE_c = \frac{\|\mathbf{u}^{gt} - \hat{\mathbf{u}}\|_2}{\|\mathbf{u}^{gt}\|_2}$.

599 For BPMRI-Net, the scaling coefficient c in the spectral normalization in Eqn. (2.15) is
 600 set to 0.97. The radius r of projection Π_{B_r} on the output of **U-Subnet** is set to 1, i.e., the
 601 magnitude of all MR images was normalized to $[0, 1]$. The projection radius on the output of
 602 **S-Subnet** is set to $1/N_c$, i.e., the sensitivity maps are normalized to satisfy $\sum_l \mathbf{s}_l^* \odot \mathbf{s}_l = 1$.
 603 We adopt the BPMRI-Net with shared parameters in 10 iterations as the default setting in
 604 the following experiments.

605 **4.2. Experimental results.** We compare our BPMRI-Net with six traditional model-based
 606 PI methods including GRAPPA [19], SPIRiT [35], AC-LORAKS [20], Fast-JTV [10], EN-
 607 LIVE [24], and LINDBERG [57]. GRAPPA, SPIRiT, and AC-LORAKS are auto-calibration
 608 methods. Fast-JTV and LINDBERG methods are calibration-less methods solving recon-
 609 struction models with a joint total variation, wavelet transform-based, and dictionary-based
 610 joint sparsity regularization on images of multi-coils. All of these five compared methods
 611 reconstruct real-valued MR images. ENLIVE estimates coil profiles and the MR image simu-
 612 taneously by solving a nonlinear reconstruction model with predefined regularization terms.
 613 We also compare our method with six deep learning methods of Learning VN [23], MoDL [1],
 614 DC-CNN [46], Deep-SLR [42], OT-cycleGAN [39], and E2E-VarNet [50]. Different from our

Table 1: Comparison of average reconstruction accuracy on the testing knee dataset with 1D random masks at different sampling rates. The top and bottom groups of these compared methods are respectively the traditional methods and deep learning-based methods.

Method	$R_A = 4$				$R_A = 6$				$R_A = 8$			
	PSNR (dB)	nRMSE	SSIM	nRMSE _c	PSNR (dB)	nRMSE	SSIM	nRMSE _c	PSNR (dB)	nRMSE	SSIM	nRMSE _c
Zero-filled	27.4755	0.1684	0.7235	0.1837	26.7098	0.1840	0.6897	0.1980	26.0822	0.1977	0.6711	0.2087
GRAPPA	31.5265	0.1072	0.7865	--	29.3383	0.1369	0.7229	--	27.4601	0.1691	0.6739	--
SPIRiT	32.3176	0.0981	0.8235	--	29.5908	0.1331	0.7392	--	27.7948	0.1632	0.6940	--
AC-LORAKS	32.8592	0.0882	0.8262	--	30.8367	0.1166	0.7443	--	28.7277	0.1475	0.6962	--
Fast-JTV	33.1288	0.0895	0.8322	--	30.6359	0.1188	0.7494	--	28.7174	0.1476	0.7046	--
ENLIVE	33.1474	0.0917	0.8201	0.1201	30.4342	0.1190	0.7470	0.1349	28.2173	0.1608	0.7028	0.1938
LINDBERG	33.9495	0.0824	0.8516	--	31.0982	0.1016	0.8203	--	29.0735	0.1404	0.7480	--
OT-cycleGAN	32.2812	0.1001	0.8028	--	30.7614	0.1230	0.7716	--	29.0998	0.1404	0.7637	--
OT-cycleGAN _{su}	32.6019	0.0923	0.8396	--	31.5072	0.1046	0.8144	--	29.1005	0.1403	0.7719	--
DC-CNN	35.2884	0.0701	0.8996	--	31.6836	0.1049	0.8272	--	29.5139	0.1339	0.7760	--
Deep-SLR	34.8320	0.0737	0.8938	--	31.0806	0.1120	0.8146	--	29.1012	0.1402	0.7674	--
Learning VN	35.2619	0.0732	0.8898	0.0860	32.4121	0.0949	0.8324	0.1162	31.0124	0.1170	0.7895	0.1281
MoDL	35.2754	0.0720	0.8893	0.0859	32.6168	0.0960	0.8345	0.1099	31.2190	0.1123	0.7982	0.1252
E2E-VarNet	35.6539	0.0680	0.9006	0.0820	32.6910	0.0943	0.8418	0.1081	31.5300	0.1074	0.8156	0.1205
BPMRI-Net _C	35.3064	0.0699	0.8998	0.0846	32.6274	0.0956	0.8356	0.1083	31.3376	0.1112	0.7995	0.1250
BPMRI-Net	36.1420	0.0643	0.9129	0.0776	33.1549	0.0899	0.8536	0.1043	31.8482	0.1039	0.8268	0.1181

method, Learning VN and MoDL are two explicit-calibration methods, with inputs of the under-sampled k -space data and pre-estimated sensitivity information of multiple channels, and they output the final reconstructed complex-valued MR image. DC-CNN, Deep-SLR, OT-cycleGAN, and E2E-VarNet are calibration-less methods. Among these, DC-CNN, Deep-SLR, and OT-cycleGAN are coil-by-coil reconstruction methods that generate real-valued MR images. DC-CNN is a deep cascade of convolutional neural networks to reconstruct multiple sequences of MR images from the under-sampled data. We use DC-CNN to reconstruct multi-channel sensitivity encoded MR images for PI. Deep-SLR exploits the deep network to model the annihilation relations from the Fourier samples for low-rank matrix completion. OT-cycleGAN uses the cycle-consistent loss and Wasserstein GAN loss to train a modified cycleGAN based on unpaired training data. The supervised learning method (i.e., based on paired training data) using the network structure of OT-cycleGAN is named OT-cycleGAN_{su}. All comparisons are based on the public codes provided by the authors, except OT-cycleGAN which is implemented by ourselves. The hyper-parameters in the traditional methods are searched and adjusted around the parameters recommended by the authors based on the validation datasets. SPIRiT uses Tikhonov regularization with the parameter 1e-5. In ENLIVE, the regularization term is composed of the penalty term of the high Fourier coefficients of the coil sensitivities and the squared ℓ_2 -regularization term of the MR image. The regularization parameter is taken as $\alpha^{(k)} = \alpha^{(0)}q^{k-1}$ where $\alpha^{(0)} = 1$ and $q = \frac{1}{3}$. The matrix rank value in AC-LORAKS is set to 600. The joint TV and wavelet transform-based sparsity regularization parameters in Fast-JTV are taken as 1e-3 and 1e-4, and the data fitting and sparsity weights in LINDBERG are selected as 0.03 and 0.04, respectively. The parameters of the deep networks were retrained based on our MRI data. We set the number of convolution blocks in each iteration of MoDL, Deep-SLR, and DC-CNN to 10, 5, and 10, respectively, and set the iteration times of the networks to 10. The iteration number of Learning VN is set to 17, and the number of feature channels is 64.

Table 1 reports the quantitative accuracy of different reconstruction methods using the

Table 2: Comparison of average reconstruction accuracy on the testing knee dataset with 1D uniform masks at different sampling rate. The top and bottom groups of these compared methods are respectively the traditional methods and deep learning-based methods.

Method	$R_A = 4$				$R_A = 6$				$R_A = 8$			
	PSNR (dB)	nRMSE	SSIM	nRMSE _c	PSNR (dB)	nRMSE	SSIM	nRMSE _c	PSNR (dB)	nRMSE	SSIM	nRMSE _c
Zero-filled	27.6904	0.1645	0.7307	0.1806	26.4246	0.1900	0.6862	0.2020	26.3796	0.1910	0.6790	0.2043
GRAPPA	33.5378	0.0881	0.8246	--	28.9519	0.1430	0.7173	--	28.3118	0.1537	0.6952	--
SPIRiT	34.1972	0.0800	0.8600	--	29.7253	0.1315	0.7541	--	28.5998	0.1491	0.7113	--
AC-LORAKS	34.9209	0.0772	0.8474	--	30.8376	0.1169	0.7621	--	29.8489	0.1302	0.7205	--
Fast-JTV	33.8331	0.0830	0.8412	--	30.1546	0.1253	0.7519	--	29.4333	0.1362	0.7189	--
ENLIVE	33.7301	0.0875	0.8284	0.1184	29.6712	0.1384	0.7532	0.1576	28.8293	0.1410	0.7146	0.1722
LINDBERG	33.5975	0.0860	0.8420	--	30.8863	0.1149	0.8115	--	29.7463	0.1160	0.7757	--
OT-cycleGAN	34.1832	0.0809	0.8613	--	30.7972	0.1153	0.8107	--	29.8972	0.1301	0.7787	--
OT-cycleGAN _{su}	34.8543	0.0798	0.8671	--	30.9457	0.1111	0.8136	--	30.0553	0.1296	0.7791	--
DC-CNN	35.7377	0.0668	0.9048	--	31.4827	0.1071	0.8254	--	30.1534	0.1246	0.7903	--
Deep-SLR	35.1925	0.0709	0.8978	--	30.9542	0.1136	0.8158	--	30.1272	0.1280	0.7810	--
Learning VN	36.1949	0.0648	0.8995	0.0798	32.6385	0.0985	0.8401	0.1103	31.3205	0.1088	0.8098	0.1239
MoDL	36.1948	0.0657	0.8996	0.0799	32.9120	0.0926	0.8410	0.1065	31.6014	0.1067	0.8106	0.1210
E2E-VarNet	36.2462	0.0636	0.9076	0.0776	33.3570	0.0876	0.8570	0.1020	32.1031	0.1005	0.8257	0.1146
BPMRI-Net _C	36.1952	0.0649	0.9004	0.0797	32.9991	0.0879	0.8506	0.1047	32.0418	0.1044	0.8164	0.1194
BPMRI-Net	36.6887	0.0604	0.9169	0.0742	33.5821	0.0857	0.8639	0.1003	32.2647	0.0989	0.8317	0.1137

sampled k -space data with 1D Cartesian random masks and acceleration factors of $R_A = 4, 6$, and 8 respectively on the knee dataset. Compared with model-based methods of GRAPPA, SPIRiT, AC-LORAKS, Fast-JTV, ENLIVE, and LINDBERG, our proposed method achieves better reconstruction accuracy in three measures for all three acceleration factors. Specifically, the performance of our proposed network exceeds the performance of six traditional methods by at least 2 dB. Moreover, the SSIM measures of reconstruction results of BPMRI-Net under $R_A = 6$ and $R_A = 8$ are better than the compared traditional methods under $R_A = 4$ and $R_A = 6$ respectively. The traditional methods of Fast-JTV, ENLIVE, LINDBERG, and AC-LORAKS produce higher accuracy on average compared with GRAPPA and SPIRiT using the random masks with all three sampling rates. With acceleration factors of 6 and 8, the methods of Fast-JTV, LINDBERG, and AC-LORAKS achieve better reconstruction performance than ENLIVE. Among them, LINDBERG learns an adaptive joint-sparse representation by updating the dictionary and obtains optimal reconstruction accuracy using the Cartesian random masks at three sampling rates. In the compared deep learning methods, methods of DC-CNN, Deep-SLR, Learning VN, MoDL, and E2E-VarNet incorporate the MR imaging mechanism into the network structures and reconstructed images of higher quality, compared with OT-cycleGAN_{su}. Compared with these deep learning methods, the proposed method achieves the best reconstruction performance. On average, BPMRI-Net outperforms OT-cycleGAN_{su}, DC-CNN, and Deep-SLR by at least 1.5 dB. Compared with the methods of Learning VN and MoDL that generate final MR images using pre-estimated sensitivity maps, E2E-VarNet improves the reconstruction results in nRMSE_c over the complex-valued images, by estimating the coil profiles through the deep network. Our method jointly updates the sensitivity maps and the MR image with network iterations, achieving better performance. Compared with BPMRI-Net_C, i.e., our method in the conference version [37], the reconstruction accuracy of BPMRI-Net is better on average, showing that the network proposed in this journal paper is more effective.

Table 2 reports the accuracy of different reconstruction methods using the sampled k -

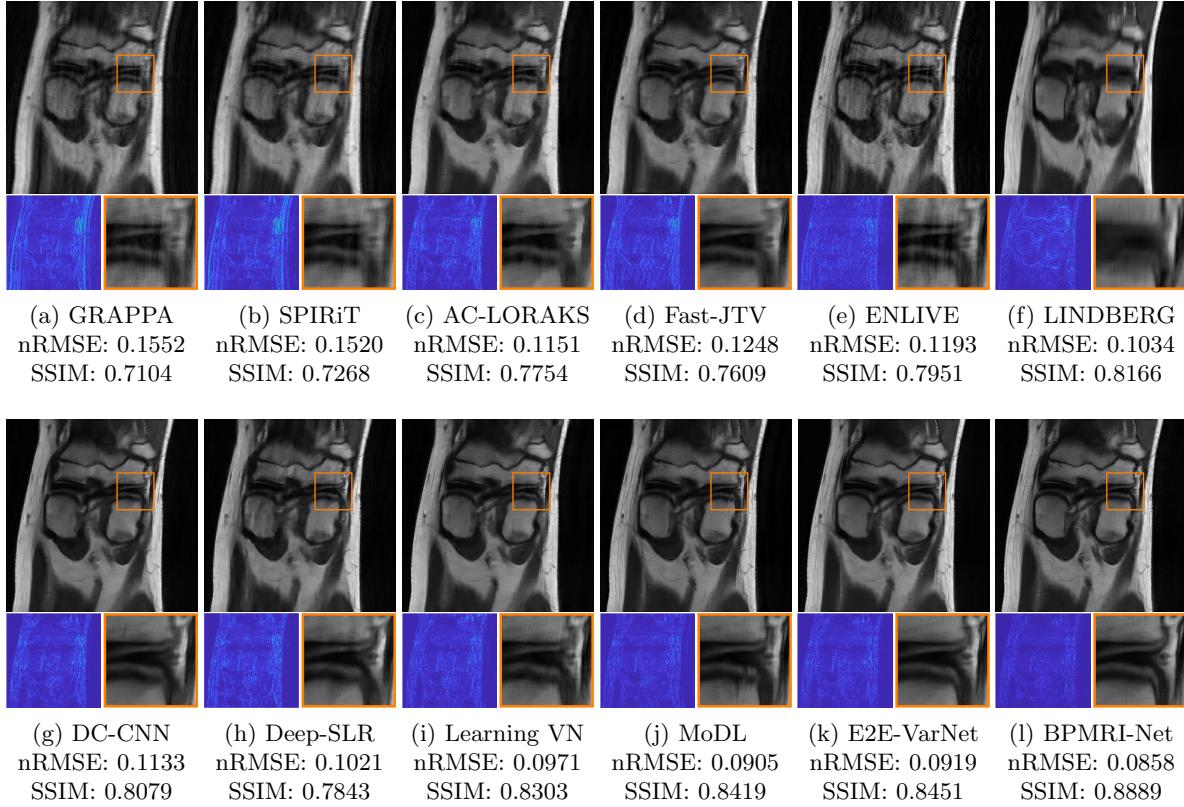


Figure 4: Examples of reconstructed knee MR images using a random mask at $6 \times$ acceleration.

space data with 1D Cartesian uniform masks and acceleration factors of $R_A = 4, 6$, and 8 respectively on the knee dataset. Similar to the results in Table 1, compared with six traditional model-based PI methods and seven deep learning methods, the accuracy of the proposed method is the best in all three evaluation measures. In addition, compared with the reconstruction results using the random masks in Table 1, the SPIRiT and AC-LORAKS methods achieved better reconstruction accuracy using the uniform masks in Table 2. In particular, the PSNR value of reconstructed images in AC-LORAKS is higher than that in LINDBERG using the acceleration factors of $R_A = 4$ and 8. With acceleration factors of 4 and 6, the reconstruction accuracy of SPIRiT is higher than that of ENLIVE. In the case of $R_A = 4$, the performance of SPIRiT is close to that of OT-cycleGAN.

Besides the quantitative evaluation of these methods, the visual comparisons of reconstructed images using random masks at $6 \times$ and $8 \times$ acceleration are shown in Figure 4 and Figure 5. The method proposed in this paper produces better images with high quality, which have no obvious artifacts and contain richer details. There are obvious vertical stripe artifacts in the results reconstructed by traditional methods GRAPPA, SPIRiT, and ENLIVE. The reconstructed images of AC-LORAKS are not smooth enough, while the reconstructed results of LINDBERG are too smooth. For a further comparison, we provide additional abso-

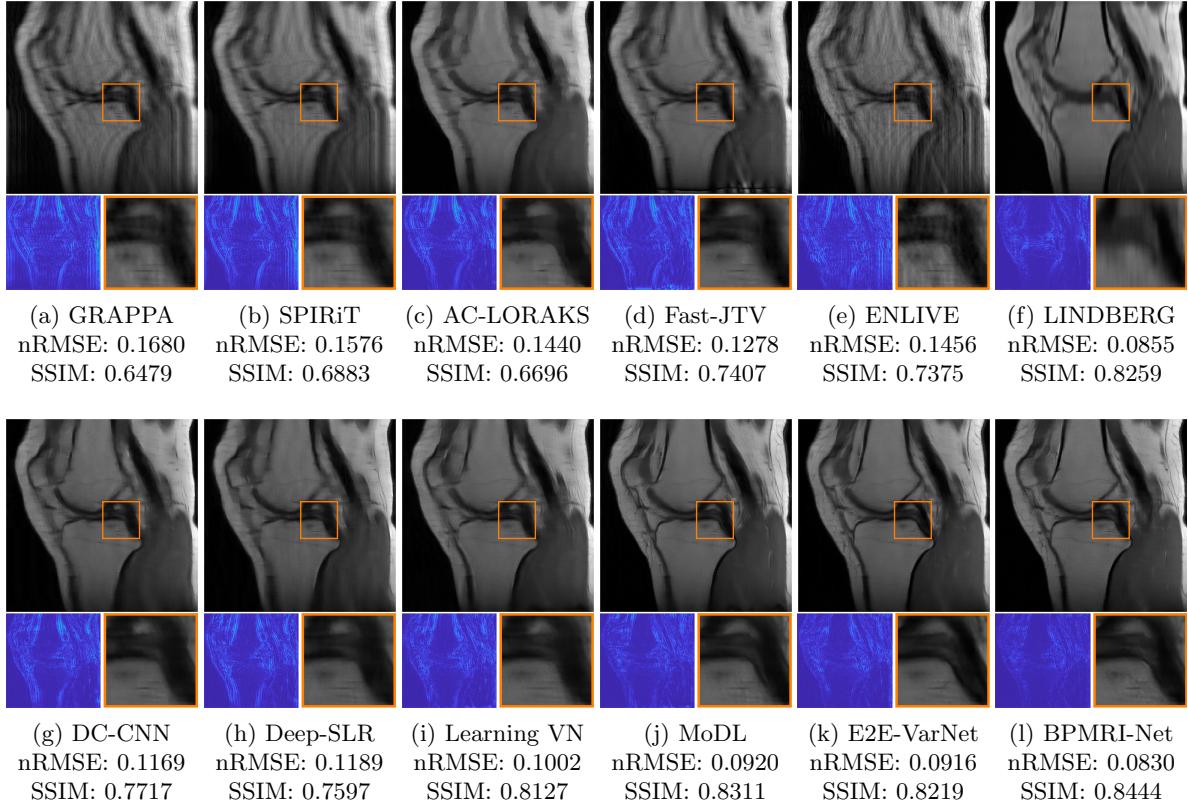


Figure 5: Examples of reconstructed knee MR images using a random mask at 8 × acceleration.

686 lute error maps relative to the ground-truth image and zoom in on a part of the reconstructed
 687 image. The reconstruction of Fast-JTV, DC-CNN, and Deep-SLR are slightly fuzzy or have
 688 slight artifacts. E2E-VarNet uses full end-to-end learning similar to our proposed method and
 689 does a good job of image reconstruction. For a higher acceleration factor in the Figure 5, our
 690 proposed network can still reconstruct high-quality images with accurate anatomical struc-
 691 tures. All the reconstruction results of the six traditional methods have serious artifacts, and
 692 their reconstruction accuracy values are not satisfactory. The reconstructed images of deep
 693 learning methods of DC-CNN, Deep-SLR, and Learning VN are too smooth to lack some
 694 details and have a few artifacts.

695 **4.3. Effect of different number of ACS lines.** Figure 6 shows the PSNR values of re-
 696 constructions by different methods using random masks of the different number of ACS lines,
 697 at 4× and 6 × acceleration factors respectively. When the number of ACS lines is small
 698 (i.e., the line number is less than 10), the pre-estimated sensitivity maps or interpolation
 699 kernels using these ACS lines are not accurate, which will lead to a decline in the quality
 700 of the final reconstructed images. We can overcome this limitation by estimating sensitiv-
 701 ity maps with deep networks. For example, BPMRI-Net and E2E-VarNet methods achieved

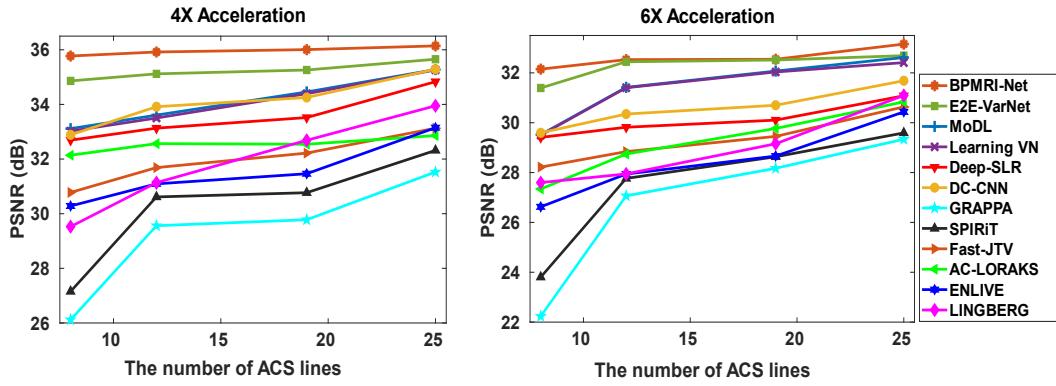


Figure 6: Effect of different number of ACS lines on reconstruction quality on the testing knee dataset using $4\times$ and $6\times$ random masks.

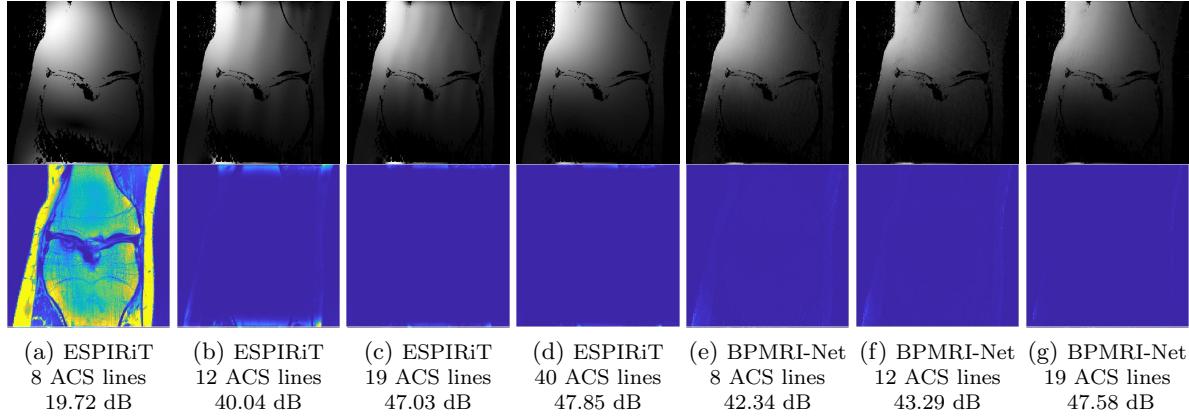


Figure 7: Examples of reconstructed sensitivity maps using $4\times$ random masks. The first row shows the estimated sensitivity maps. (a)-(d) are the pre-estimated sensitivities using ESPIRiT; (e)-(g) are the sensitivities reconstructed by BPMRI-Net. All maps are computed by multiplying the estimated sensitivity maps with a mask that is based on the image's amplitude and a threshold of $1e-4$. The second row displays the residual signal maps in the null space. The residual signal maps and the PSNR values are calculated based on the original fully-sampled coil images and the reconstructed coil images $\{\hat{v}_l\}_{l=1}^{N_c}$ using the estimated sensitivity maps $\{\hat{s}_l\}_{l=1}^{N_c}$ by $\hat{v}_l = \text{diag}(\hat{s}_l) \sum_{i=1}^{N_c} \text{diag}(\hat{s}_i)^\top v_i$, $l \in \{1, 2, \dots, N_c\}$.

higher reconstruction accuracy. Especially when fewer low-frequency lines exist in the center of k -space (ACS), good reconstruction results can still be obtained. Since the calibration-less methods of Fast-JTV, DC-CNN, and Deep-SLR, do not estimate the sensitivity information based on ACS, the reconstruction accuracy is not seriously affected by the number of ACS lines. The explicit-calibration methods of Learning VN and MoDL use the pre-estimated sen-

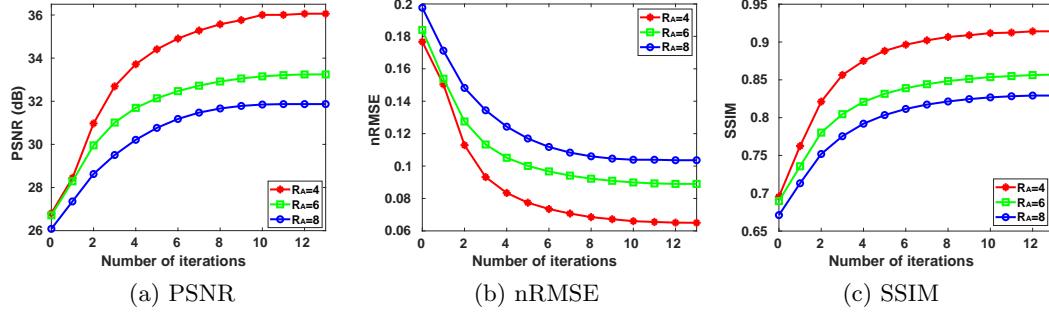


Figure 8: Average test PSNR, nRMSE, and SSIM values for increasing network iterations on the testing knee dataset using random masks at 4, 6, and 8 × acceleration.

sitivity maps, and their curves of PNSR values are under the curve of our approach across the different number of ACS lines, especially when the number of ACS lines is small. We also exhibit examples of the sensitivity maps reconstructed by our method with different numbers of sampled ACS lines at 4× acceleration in Figure 7. It can be found that the sensitivities estimated by our network at different ACS sizes are close to the sensitivity map estimated by ESPIRiT using the sufficient sampling of k -space center (e.g., 40 ACS lines). In addition, the Figure 7 shows the residual images in the null space of the linear transformation $(\text{diag}(\mathbf{s}_l) \sum_{i=1}^{N_c} \text{diag}(\mathbf{s}_i)^\top - \mathbf{I})$, $l \in \{1, 2, \dots, N_c\}$ [56] and the accuracy (i.e., the PSNR value) of the reconstructed coil images $\{\hat{\mathbf{v}}_l\}_{l=1}^{N_c}$ using the estimated sensitivity maps $\{\hat{\mathbf{s}}_l\}_{l=1}^{N_c}$ by $\hat{\mathbf{v}}_l = \text{diag}(\hat{\mathbf{s}}_l) \sum_{i=1}^{N_c} \text{diag}(\hat{\mathbf{s}}_i)^\top \mathbf{v}_i$, $l \in \{1, 2, \dots, N_c\}$, where $\text{diag}(\cdot)$ is the operation that transforms a vector into a diagonal matrix and \mathbf{I} is an identity matrix. The sensitivities estimated by the proposed network yield smaller residual signals and higher reconstruction accuracy compared with the pre-estimated sensitivities, particularly when fewer ACS lines are used (e.g., 8 ACS lines). Note that our network does not have supervision on coil sensitivity during training.

4.4. Convergence validation. To experimentally validate the convergence property of BPMRI-Net and the effect of the number of iterations (i.e., K) on reconstruction performance, we plotted the reconstruction precision (i.e., PSNR, nRMSE, and SSIM values) of the MR image as functions of the iteration number at three different acceleration factors in Figure 8. It qualitatively shows that BPMRI-Net, as an iterative network, exhibits stable and fast convergence, and the accuracy of reconstruction gradually increases with the increase of iterations. Specifically, the average PSNR, nRMSE, and SSIM values of the MR image converge at around 10 iterations. Visually, as shown in Figure 9, BPMRI-Net sequentially outputs MR images from iteration 0 to iteration 14 with increasingly clearer images and smaller errors, and the accuracy of the output images tends to be stable.

We next validate some of the assumptions and conditions for the convergence analysis. We first experimentally examine whether the learned sub-networks $F_u = Id + \tilde{F}_u$ and $F_s = Id + \tilde{F}_s$ satisfy Assumption 3.6, i.e., $0 < L < 1$. Given a test sample, we estimate the Lipschitz

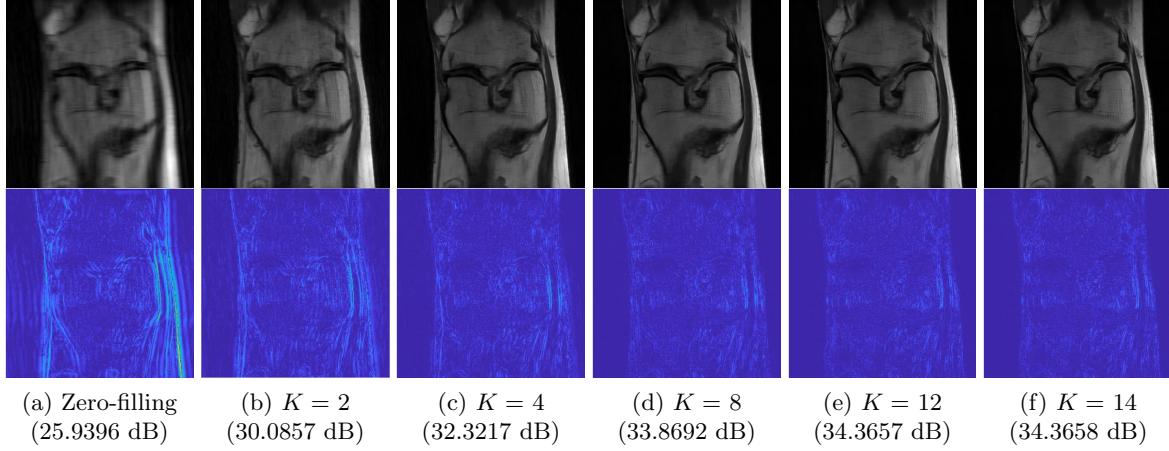


Figure 9: Illustration of reconstructed MR images by BPMRI-Net in different iterations (i.e., K) and corresponding absolute error maps, using a random mask at $6 \times$ acceleration.

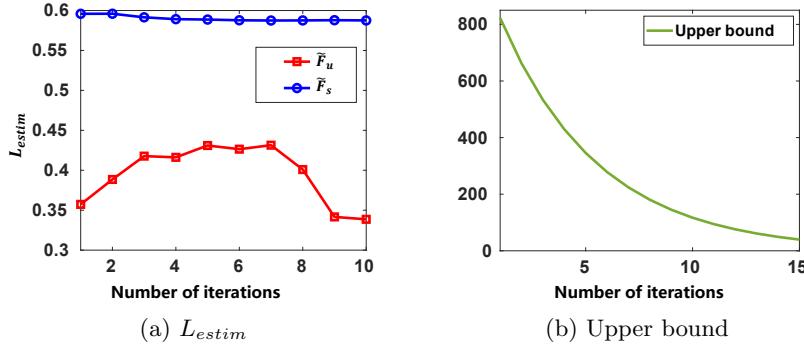


Figure 10: (a) Demonstration of the changes in estimates of the Lipschitzian constant with iterations for two subnetworks using a knee data with a $4 \times$ random mask. (b) Illustration of the upper bound of convergence error in different iterations in [Theorem 3.16](#).

735 constant L_{estim} for each iteration of the sub-networks, based on

$$\frac{\|\tilde{F}_u(\tilde{\mathbf{u}}^{(k)}) - \tilde{F}_u(\tilde{\mathbf{u}}^{(k)} + \mathbf{e}_u^{(k)})\|}{\|\mathbf{e}_u^{(k)}\|}, \quad \frac{\|\tilde{F}_s(\tilde{\mathbf{s}}_l^{(k)}) - \tilde{F}_s(\tilde{\mathbf{s}}_l^{(k)} + \mathbf{e}_s^{(k)})\|}{\|\mathbf{e}_s^{(k)}\|},$$

736 where the elements in $\mathbf{e}_u^{(k)}$ and $\mathbf{e}_s^{(k)}$ are sampled from a Normal distribution. As shown
737 in [Figure 10\(a\)](#), the estimated Lipschitz constant of two sub-networks for 10 iterations satisfies
738 $0 < L_{estim} < 1$, especially for later iterations. Then we show the estimated upper bound of the
739 convergence error in different iterations using the Eqn. (3.12) of [Theorem 3.16](#) in [Figure 10\(b\)](#).
740

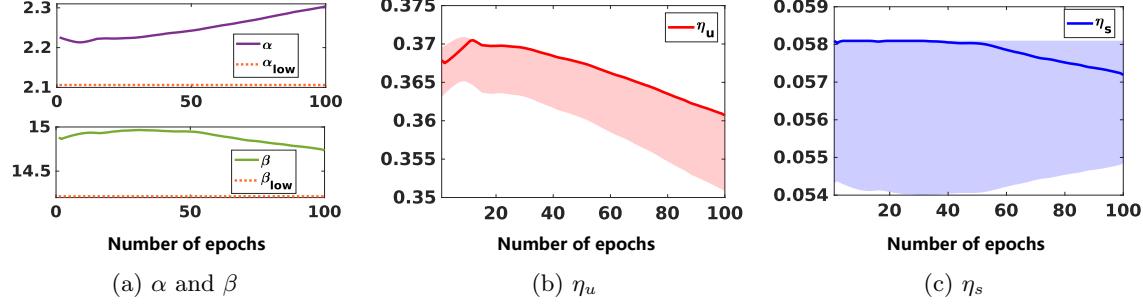


Figure 11: Illustration of learned parameters α , β , η_u and η_s on a training knee dataset with a $4 \times$ random mask in different training epochs.

We can find that the estimated error exponentially decays with iterations. Finally, we show the parameters α and β along with the training process in Figure 11(a). All parameters are larger than the estimated lower bounds (i.e., two dotted lines in the figure), which is consistent with Condition 3.8. The Lipschitz constant L in the lower bounds is based on a pre-estimated value (i.e., 0.97^{10} or 0.97^3). Figure 11(b) and (c) show the learned parameters η_u and η_s and indicate the range (i.e., the shaded parts of the figure) in which these parameters need to be satisfied. We can see that these parameters satisfy the Condition 3.10.

Table 3: Performance comparison on the testing knee dataset for different network settings.

Accel.	Method	Using image update block?	Using sensitivity update block?	PSNR (dB)
$R_A = 4$	Net-NoS	yes	no	34.8072
	Net-NoU	no	yes	32.5479
	BPMRI-Net	yes	yes	36.142
$R_A = 6$	Net-NoS	yes	no	31.9816
	Net-NoU	no	yes	29.76
	BPMRI-Net	yes	yes	33.1549
$R_A = 8$	Net-NoS	yes	no	30.1394
	Net-NoU	no	yes	27.9723
	BPMRI-Net	yes	yes	31.8482

4.5. Ablation study. To test the effectiveness of two sub-networks for sensitivity and image updating, we respectively set **S-Subnet** (i.e., F_s) and **U-Subnet** (i.e., F_u) as an identity mapping without learning, and the corresponding networks are denoted as Net-NoS and Net-NoU. The performance comparison of trained Net-NoS, Net-NoU, and BPMRI-Net using 1D random masks at $4 \times$, $6 \times$, and $8 \times$ accelerations is shown in Table 3. We can find that both sub-networks are beneficial for improving the reconstruction quality. In addition, the MR image update subnetwork can improve the accuracy of the algorithm to a greater extent.

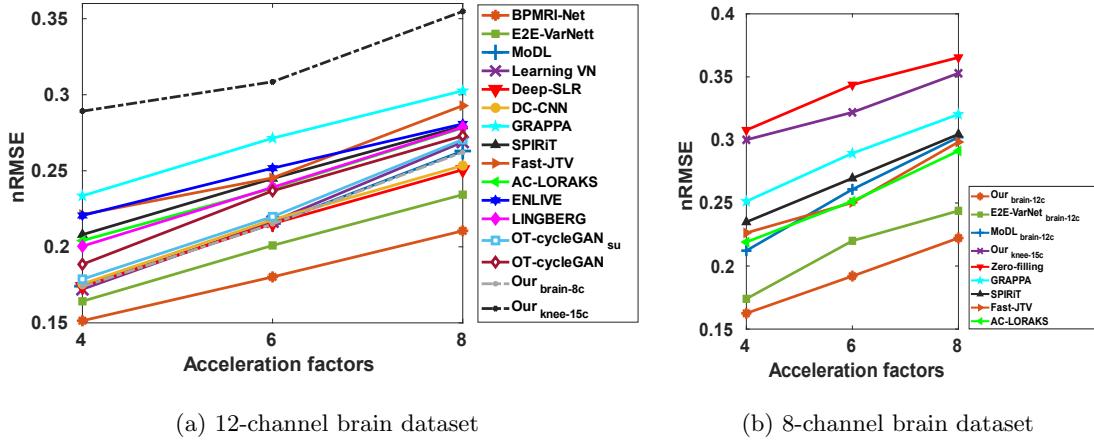


Figure 12: Comparison of average nRMSE on the testing brain datasets with 1D random masks at different sampling rates.

4.6. The performance on the brain data. We also experimentally evaluated the performance of the proposed method on our collected brain datasets. Figure 12(a) shows the average reconstruction accuracy (i.e., nRMSE) of different methods on the 12-channel brain test dataset using $4 \times$, $6 \times$, and $8 \times$ Cartesian random masks. As shown in Figure 12(a), the proposed BPMRI-Net achieves the lowest reconstruction error at all three sampling rates. The reconstruction results of the compared deep learning methods are on average better than those of the traditional model-based methods. In particular, DC-CNN, Deep-SLR, and E2E-VarNet achieve higher reconstruction accuracy than Learning VN, MoDL, and OT-cycleGAN, especially at the higher acceleration factor of 8.

Figure 13 visually compares our BPMRI-Net with other 11 reconstruction methods using a $4 \times$ Cartesian random mask on the 12-channel brain data. Compared with GRAPPA, SPIRiT, AC-LORAKS, Fast-JTV, and ENLIVE, the deep learning methods can restore the basic anatomical structure of the brain without serious artifacts. In particular, our proposed method obtains the smallest reconstruction error and reconstructs more clear image details.

In addition, we verify the generalization ability of our proposed method for different coil configurations. We respectively test the performance of Our_{knee-15c} and Our_{brain-12c} on the 8-channel brain dataset using $4 \times$, $6 \times$, and $8 \times$ Cartesian random masks. Our_{knee-15c} is the proposed network trained on a 15-channel knee dataset and Our_{brain-12c} is the network trained on a 12-channel brain dataset. The 8-channel brain and 12-channel brain datasets are derived from the same MRI system under different coil settings. The 8-channel brain and 15-channel knee datasets are derived from different coil settings of different MR imaging systems. As shown in Figure 12(b), the reconstruction performance of our network trained on the 12-channel brain dataset (i.e., Our_{brain-12c}) can outperform the reconstruction performance of traditional methods of GRAPPA, SPIRiT, Fast-JTV, and AC-LORAKS, and deep learning methods MoDL and E2E-VarNet trained on the same 12-channel brain dataset, at three

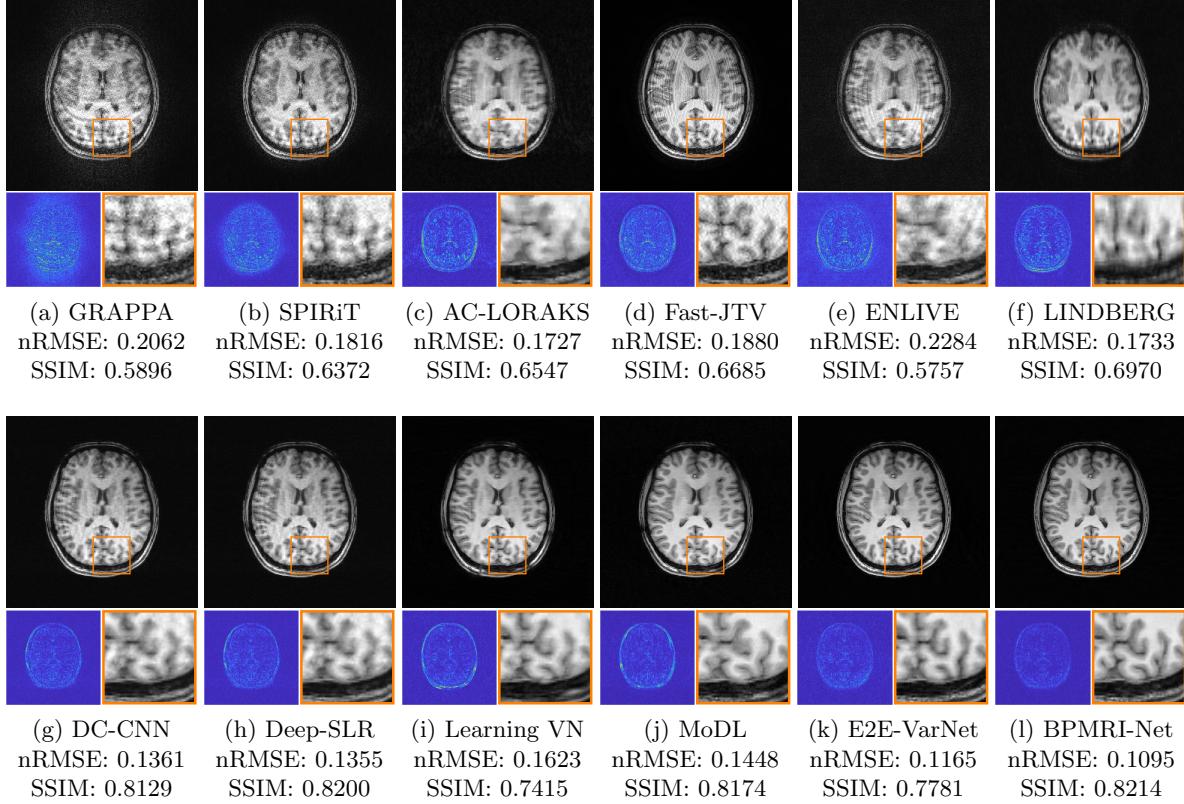


Figure 13: Examples of reconstructed brain images using a random mask at 4 × acceleration.

781 sampling rates. The reconstruction accuracy of our network trained on the 15-channel knee
 782 dataset (i.e., Our_{knee-15c}) is lower than that of the four traditional methods. Furthermore,
 783 we test the reconstruction performance of our proposed network on the 12-channel brain test
 784 dataset, which is trained on the 8-channel brain dataset (denoted as Our_{brain-8c}) or the 15-
 785 channel knee dataset (denoted as Our_{knee-15c}). As shown in Figure 12(a), the reconstruction
 786 accuracy of Our_{brain-8c} is higher than that of the six compared traditional methods and close
 787 to that of the MoDL method, which is trained on the 12-channel brain data. Similarly, the
 788 reconstruction performance of the six traditional methods is better than that of Our_{knee-15c}.

789 **5. Conclusion and discussion.** In this work, we proposed a novel reconstruction model,
 790 its optimization algorithm, and the unrolled deep network to simultaneously estimate the MR
 791 image and sensitivity maps for parallel imaging from sampled k -space data of multi-coils.
 792 After building a reconstruction model coupling undetermined regularization terms of MR
 793 image and sensitivity maps, we unfold its optimization algorithm to be a deep architecture, in
 794 which two regularization terms are implicitly learned by substituting corresponding proximal
 795 operators using two deep convolutional sub-networks. Then we analyzed the convergence of
 796 the proposed network as an iterative algorithm under mild assumptions which can be satisfied

797 in our network. Finally, we validate the algorithm performance and the proposed theory with
 798 experiments. To the best of our knowledge, this is the first work to propose a deep architecture
 799 to implicitly learn both regularizations of the MR image and sensitivity maps for parallel
 800 imaging in a blind setting with convergence analysis of the proposed network. The proposed
 801 reconstruction model is an optimization model of the sum of a data term represented by a
 802 coupled smooth function and regularization terms represented by non-convex and non-smooth
 803 functions, which is a general model. Therefore, as a generalized framework, our network can
 804 be generically applied to other blind inverse problems such as image deblurring and blind
 805 image super-resolution.

806 There are two limitations of our method that should be addressed in future research.
 807 Firstly, compared with the non-learning-based methods, our proposed BPMRI-Net relies on
 808 training data to learn hyper-parameters and proximal operators of regularizers of the MR
 809 images and sensitivities for parallel MR image reconstruction. If the training and test data
 810 come from different coil configurations, especially using different MRI systems, the learned
 811 network from the training data may have limited generalization ability to the test data.
 812 Secondly, our convergence analysis is a convergence proof in a weak form that the sequence
 813 generated by BPMRI-Net is convergent, but does not guarantee that the sequences with
 814 different initialization converge to the same fixed point.

815 In the future work, we are interested in designing a learning-based reconstruction method,
 816 which have better generalization ability to different coil configurations and MRI systems.
 817 We plan to integrate the hyper-parameters of the coil configurations in imaging systems into
 818 the proposed network structure and train a reconstruction network adaptive to different coil
 819 configurations or imaging systems. Furthermore, we plan to design a learning-based recon-
 820 struction method that can not only reconstruct MR images in high precision but also estimate
 821 the uncertainty of the reconstruction, which can provide reliable and explainable results for
 822 subsequent analysis tasks. We are also interested in theoretical understanding of the proposed
 823 model-driven learning-based method, compared with the traditional regularization-based and
 824 deep learning-based models, from the perspective of approximation and generalization bounds
 825 based on statistical machine learning.

826 **Appendix A. Proof of Lemma 3.12.**

827 *Proof.* First we prove the strong convexity property of $H(\mathbf{u}, \mathbf{S})$ w.r.t. \mathbf{s}_l and \mathbf{u} , respec-
 828 tively. From the definition of $H(\mathbf{u}, \mathbf{S})$ in Eqn. (2.2), we obtain

$$829 \quad (A.1) \quad \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}) = \text{diag}(\mathbf{u})^\top \mathbf{A}^\top (\mathbf{A} \text{diag}(\mathbf{s}_l) \mathbf{u} - \mathbf{y}_l) + \beta \mathbf{s}_l,$$

$$832 \quad (A.2) \quad \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}) = \sum_l \text{diag}(\mathbf{s}_l)^\top \mathbf{A}^\top (\mathbf{A} \text{diag}(\mathbf{u}) \mathbf{s}_l - \mathbf{y}_l) + \alpha \mathbf{u},$$

$$835 \quad (A.3) \quad \nabla_{\mathbf{s}_l^\top} \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}) = \text{diag}(\mathbf{u})^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{u}) + \beta \mathbf{I},$$

$$838 \quad (A.4) \quad \nabla_{\mathbf{u}^\top} \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}) = \sum_l \text{diag}(\mathbf{s}_l)^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{s}_l) + \alpha \mathbf{I},$$

where the superscript \top represents conjugate transpose and $\text{diag}(\cdot)$ is the operation that transforms a vector into a diagonal matrix. $\nabla_{\mathbf{s}_l^\top} \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S})$ and $\nabla_{\mathbf{u}^\top} \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S})$ are respectively the Hessian matrix of the function $H(\mathbf{u}, \mathbf{S})$ w.r.t. \mathbf{s}_l and \mathbf{u} , i.e., $(\nabla_{\mathbf{s}_l^\top} \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}))_{i,j} = \frac{\partial^2 H}{\partial \mathbf{s}_{l,i} \partial \mathbf{s}_{l,j}}$, $(\nabla_{\mathbf{u}^\top} \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}))_{i,j} = \frac{\partial^2 H}{\partial \mathbf{u}_i \partial \mathbf{u}_j^*}$, where the superscript $*$ represents the conjugate operation.

Then for all $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{x} \neq \mathbf{0}$, we have

$$\mathbf{x}^\top [\text{diag}(\mathbf{u})^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{u})] \mathbf{x} = [\text{diag}(\mathbf{u}) \mathbf{x}]^\top (\mathbf{A}^\top \mathbf{A}) [\text{diag}(\mathbf{u}) \mathbf{x}] \geq 0,$$

847

$$\mathbf{x}^\top [\text{diag}(\mathbf{s}_l)^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{s}_l)] \mathbf{x} = [\text{diag}(\mathbf{s}_l) \mathbf{x}]^\top (\mathbf{A}^\top \mathbf{A}) [\text{diag}(\mathbf{s}_l) \mathbf{x}] \geq 0.$$

850 Hence, from Eqn. (A.3) and Eqn. (A.4), $H(\mathbf{u}, \mathbf{S})$ is strongly convex w.r.t. \mathbf{s}_l with constant β
851 and strongly convex w.r.t. \mathbf{u} with α .

852 Next, we prove the Lipschitz properties of the gradients of function H . (1) For all $\mathbf{S}_1 =$
853 $\{\mathbf{s}_{l,1}\}_l$, $\mathbf{S}_2 = \{\mathbf{s}_{l,2}\}_l$, where $\mathbf{s}_{l,1}, \mathbf{s}_{l,2} \in \mathbb{C}^N$ ($l \in \{1, 2, \dots, N_c\}$), and $\mathbf{u} \in \mathbb{C}^N$, we have

$$\begin{aligned} 854 \quad \|\nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}_1) - \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}_2)\| &\stackrel{(i)}{\leq} \left\| \text{diag}(\mathbf{u})^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{u}) + \beta \mathbf{I} \right\| \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\| \\ 855 \quad &\stackrel{(ii)}{\leq} (\left\| \text{diag}(\mathbf{u})^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{u}) \right\| + \beta) \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\| \\ 856 \quad &\stackrel{(iii)}{\leq} (B_u^2 \|\mathbf{A}\|^2 + \beta) \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\|, \end{aligned}$$

858 where (i) comes from Eqn. (A.3), (ii) uses Triangle Inequality which holds in \mathbb{C}^N and (iii)
859 arises from Assumption 3.4. This completes our proof of property in Eqn. (3.5). (2) Property
860 in Eqn. (3.6) can be proved similarly. For all $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{C}^N$ and $\mathbf{s}_l \in \mathbb{C}^N$ we obtain

$$\begin{aligned} 861 \quad \|\nabla_{\mathbf{u}} H(\mathbf{u}_1, \mathbf{S}) - \nabla_{\mathbf{u}} H(\mathbf{u}_2, \mathbf{S})\| &\leq \left(\left\| \sum_l \text{diag}(\mathbf{s}_l)^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{s}_l) \right\| + \alpha \right) \|\mathbf{u}_1 - \mathbf{u}_2\| \\ 862 \quad &\leq \left(\sum_l (B_s^2 \|\mathbf{A}\|^2) + \alpha \right) \|\mathbf{u}_1 - \mathbf{u}_2\| \\ 863 \quad &\leq (N_c B_s^2 \|\mathbf{A}\|^2 + \alpha) \|\mathbf{u}_1 - \mathbf{u}_2\|. \end{aligned}$$

865 (3) From Eqn. (A.1) and Eqn. (A.2), we obtain

$$866 \quad (\text{A.5}) \quad \nabla_{\mathbf{u}^\top} \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}) = 2 \text{diag}(\mathbf{u})^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{s}_l) - \text{diag}(\mathbf{A}^\top \mathbf{y}_l),$$

868

$$869 \quad (\text{A.6}) \quad \nabla_{\mathbf{s}_l^\top} \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}) = 2 \text{diag}(\mathbf{s}_l)^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{u}) - \text{diag}(\mathbf{A}^\top \mathbf{y}_l),$$

871 where $(\nabla_{\mathbf{u}^\top} \nabla_{\mathbf{s}_l} H(\mathbf{u}, \mathbf{S}))_{i,j} = \frac{\partial^2 H}{\partial \mathbf{s}_{l,i} \partial \mathbf{u}_j^*}$ and $(\nabla_{\mathbf{s}_l^\top} \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}))_{i,j} = \frac{\partial^2 H}{\partial \mathbf{u}_i \partial \mathbf{s}_{l,j}^*}$. Then, for all $\mathbf{s}_{l,1}, \mathbf{s}_{l,2} \in$

872 \mathbb{C}^N and $\mathbf{u} \in \mathbb{C}^N$ we get

$$\begin{aligned}
873 \quad & \|\nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}_1) - \nabla_{\mathbf{u}} H(\mathbf{u}, \mathbf{S}_2)\| \leq \sum_l \|\nabla_{\mathbf{u}} H_l(\mathbf{u}, \mathbf{s}_{l,1}) - \nabla_{\mathbf{u}} H_l(\mathbf{u}, \mathbf{s}_{l,2})\| \\
874 \quad & \leq \sum_l \left\| 2\text{diag}(\mathbf{u})^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{s}_l) - \text{diag}(\mathbf{A}^\top \mathbf{y}_l) \right\| \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\| \\
875 \quad & \leq \sum_l \left(2 \|\mathbf{A}\|^2 B_s B_u + \left\| \mathbf{A}^\top \mathbf{y}_l \right\| \right) \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\| \\
876 \quad & \leq \left(2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \left\| \mathbf{A}^\top \mathbf{y}_l \right\| \right) \sum_l \|\mathbf{s}_{l,1} - \mathbf{s}_{l,2}\|,
\end{aligned}$$

877 which corresponds to Eqn. (3.7). (4) Similarly, for all $\mathbf{u}_1, \mathbf{u}_2$ and $\mathbf{s}_l \in \mathbb{C}^N$, we can derive that

$$\begin{aligned}
879 \quad & \|\nabla_{\mathbf{s}_l} H(\mathbf{u}_1, \mathbf{S}) - \nabla_{\mathbf{s}_l} H(\mathbf{u}_2, \mathbf{S})\| \leq \left\| 2\text{diag}(\mathbf{s}_l)^\top \mathbf{A}^\top \mathbf{A} \text{diag}(\mathbf{u}) - \text{diag}(\mathbf{A}^\top \mathbf{y}_l) \right\| \|\mathbf{u}_1 - \mathbf{u}_2\| \\
880 \quad & \leq \sum_l \left(2 \|\mathbf{A}\|^2 B_s B_u + \left\| \mathbf{A}^\top \mathbf{y}_l \right\| \right) \|\mathbf{u}_1 - \mathbf{u}_2\| \\
881 \quad & \leq \left(2 \|\mathbf{A}\|^2 B_s B_u + \max_{1 \leq l \leq N_c} \left\| \mathbf{A}^\top \mathbf{y}_l \right\| \right) \|\mathbf{u}_1 - \mathbf{u}_2\|,
\end{aligned} \quad \blacksquare$$

882 which results in Eqn. (3.8).

884 Appendix B. Proof of Theorem 3.16.

885 *Proof.* Taking $m, k \in \mathbb{N}$ which satisfy $m > k$, by Lemma 3.15, we have

$$\begin{aligned}
886 \quad & \left\| \mathbf{z}^{(m)} - \mathbf{z}^{(k)} \right\|_{\text{split}} \leq \left\| \mathbf{z}^{(m)} - \mathbf{z}^{(m-1)} \right\|_{\text{split}} + \left\| \mathbf{z}^{(m-1)} - \mathbf{z}^{(m-2)} \right\|_{\text{split}} + \cdots + \left\| \mathbf{z}^{(k+1)} - \mathbf{z}^{(k)} \right\|_{\text{split}} \\
887 \quad & \leq q^k \left\| \mathbf{z}^{(1)} - \mathbf{z}^{(0)} \right\|_{\text{split}} \sum_{v=0}^{m-k-1} q^v \\
888 \quad (B.1) \quad & \leq \frac{q^k}{1-q} \left\| \mathbf{z}^{(1)} - \mathbf{z}^{(0)} \right\|_{\text{split}}.
\end{aligned}$$

889 For any given $\epsilon > 0$, there exists $K' = \lfloor \log_q \frac{(1-q)\epsilon}{\left\| \mathbf{z}^{(1)} - \mathbf{z}^{(0)} \right\|_{\text{split}}} \rfloor + 1$ s.t. for all $m, k > K'$, we have

890 $\left\| \mathbf{z}^{(m)} - \mathbf{z}^{(k)} \right\|_{\text{split}} \leq \epsilon$. Hence we obtain

$$892 \quad (B.2) \quad \left\| \mathbf{z}^{(m)} - \mathbf{z}^{(k)} \right\|_2 \leq \left\| \mathbf{z}^{(m)} - \mathbf{z}^{(k)} \right\|_{\text{split}} \leq \epsilon.$$

893 This means sequence $\{\mathbf{z}^{(k)}\}$ is a Cauchy sequence. Then by the completeness of space
894 $(\mathbb{R}^{(N_c+1)N'}, \|\cdot\|_2)$, the sequence has a limit point $\hat{\mathbf{z}} = ((\hat{\mathbf{u}})^\top, (\hat{\mathbf{s}}_1)^\top, (\hat{\mathbf{s}}_2)^\top, \dots, (\hat{\mathbf{s}}_{N_c})^\top)^\top$.

895 Let $m \rightarrow +\infty$ in Eqn. (B.1), then we get that the error decays exponentially in iterations:

$$896 \quad \left\| \hat{\mathbf{z}} - \mathbf{z}^{(k)} \right\|_2 \leq \frac{q^k}{1-q} \left\| \mathbf{z}^{(1)} - \mathbf{z}^{(0)} \right\|_{\text{split}} \leq \frac{q^k}{1-q} \left(\left\| \mathbf{u}^{(1)} - \mathbf{u}^{(0)} \right\|_2 + \sum_{l=1}^{N_c} \left\| \mathbf{s}_l^{(1)} - \mathbf{s}_l^{(0)} \right\|_2 \right).$$

REFERENCES

- [1] H. K. AGGARWAL, M. P. MANI, AND M. JACOB, *MoDL: Model-based deep learning architecture for inverse problems*, IEEE Transactions on Medical Imaging, 38 (2018), pp. 394–405.
- [2] 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, 37 (2020), pp. 105–116.
- [3] M. AKÇAKAYA, T. A. BASHA, B. GODDU, L. A. GOEPFERT, K. V. KISSINGER, V. TAROKH, W. J. MANNING, AND R. NEZAFAT, *Low-dimensional-structure self-learning and thresholding: Regularization beyond compressed sensing for MRI reconstruction*, Magnetic Resonance in Medicine, 66 (2011), pp. 756–767.
- [4] M. ARVANTE, S. VISHWANATH, A. H. TEWFIK, AND J. I. TAMIR, *Deep J-Sense: Accelerated mri reconstruction via unrolled alternating optimization*, in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2021.
- [5] H. ATTTOUCH, 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, 35 (2010), pp. 438–457.
- [6] H. ATTTOUCH, P. REDONT, AND A. SOUBEYRAN, *A new class of alternating proximal minimization algorithms with costs-to-move*, SIAM Journal on Optimization, 18 (2007), pp. 1061–1081.
- [7] K. T. BLOCK, M. UECKER, AND J. FRAHM, *Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint*, Magnetic Resonance in Medicine, 57 (2007), pp. 1086–1098.
- [8] M. BLUMENTHAL, G. LUO, M. SCHILLING, M. HALTMEIER, AND M. UECKER, *Nlinv-net: Self-supervised end-2-end learning for reconstructing undersampled radial cardiac real-time data*, in Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM), 2022.
- [9] J. BOLTE, S. SABACH, AND M. TEBOULLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.
- [10] C. CHEN, Y. LI, AND J. HUANG, *Calibrationless parallel MRI with joint total variation regularization*, in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2013, pp. 106–114.
- [11] X. CHEN, J. LIU, Z. WANG, AND W. YIN, *Theoretical linear convergence of unfolded ista and its practical weights and thresholds*, in Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS), 2018.
- [12] Y. CHEN, W. HAGER, F. HUANG, D. PHAN, X. YE, AND W. YIN, *Fast algorithms for image reconstruction with application to partially parallel MR imaging*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 90–118.
- [13] Y. CHEN, H. LIU, X. YE, AND Q. ZHANG, *Learnable descent algorithm for nonsmooth nonconvex image reconstruction*, SIAM Journal on Imaging Sciences, 14 (2021), pp. 1532–1564.
- [14] Y. CHEN, T. XIAO, C. LI, Q. LIU, AND S. WANG, *Model-based convolutional de-aliasing network learning for parallel MR imaging*, in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2019, pp. 30–38.
- [15] J. Y. CHENG, J. M. PAULY, AND S. S. VASANAWALA, *Multi-channel image reconstruction with latent coils and adversarial loss*, in Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM), 2019.
- [16] R. Giryes, Y. C. Eldar, A. M. Bronstein, AND G. Sapiro, *Tradeoffs between convergence speed and reconstruction accuracy in inverse problems*, IEEE Transactions on Signal Processing, 66 (2018), pp. 1676–1690.
- [17] H. GOUK, E. FRANK, B. PFAHRINGER, AND M. CREE, *Regularisation of neural networks by enforcing lipschitz continuity*, Machine Learning, 110 (2018), pp. 393–416.
- [18] K. GREGOR AND Y. LECUN, *Learning fast approximations of sparse coding*, in Proceedings of the International Conference on International Conference on Machine Learning (ICML), 2010, p. 399–406.
- [19] M. A. GRISWOLD, P. M. JAKOB, R. M. HEIDEMANN, M. NITTKA, V. JELLUS, J. WANG, B. KIEFER, AND A. HAASE, *Generalized autocalibrating partially parallel acquisitions (grappa)*, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 47

- 950 (2002), pp. 1202–1210.
951 [20] J. P. HALDAR, *Autocalibrated loraks for fast constrained MRI reconstruction*, in Proceedings of the IEEE
952 International Symposium on Biomedical Imaging (ISBI), 2015, pp. 910–913.
953 [21] J. P. HALDAR AND K. SETSOMPOP, *Linear predictability in magnetic resonance imaging reconstruction:
954 Leveraging shift-invariant fourier structure for faster and better imaging*, IEEE Signal Processing
955 Magazine, 37 (2020), pp. 69–82.
956 [22] J. P. HALDAR AND J. ZHUO, *P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel
957 imaging data*, Magnetic Resonance in Medicine, 75 (2016), pp. 1499–1514.
958 [23] K. HAMMERNIK, T. KLATZER, E. KOBLER, M. P. RECHT, D. K. SODICKSON, T. POCK, AND F. KNOLL,
959 *Learning a variational network for reconstruction of accelerated MRI data*, Magnetic Resonance in
960 Medicine, 79 (2018), pp. 3055–3071.
961 [24] H. C. M. HOLME, S. ROSENZWEIG, F. ONG, R. N. WILKE, M. LUSTIG, AND M. UECKER, *ENLIVE: An
962 efficient nonlinear method for calibrationless and robust parallel imaging*, Scientific Reports, 9 (2019),
963 p. 3034.
964 [25] P. M. JAKOB, M. A. GRISOWLD, R. R. EDELMAN, AND D. K. SODICKSON, *AUTO-SMASH: A self-
965 calibrating technique for SMASH imaging*, Magnetic Resonance Materials in Physics, Biology and
966 Medicine, 7 (1998), pp. 42–54.
967 [26] K. H. JIN, D. LEE, AND J. C. YE, *A general framework for compressed sensing and parallel MRI using
968 annihilating filter based low-rank hankel matrix*, IEEE Transactions on Computational Imaging, 2
969 (2016), p. 16.
970 [27] F. KNOLL, C. CLASON, K. BREDIES, M. UECKER, AND R. STOLLBERGER, *Parallel imaging with nonlinear
971 reconstruction using variational penalties*, Magnetic Resonance in Medicine, 67 (2012), pp. 34–41.
972 [28] K. KWON, D. KIM, AND H. PARK, *A parallel mr imaging method using multilayer perceptron*, Medical
973 Physics, 44 (2017), pp. 6209–6224.
974 [29] Y. LI, R. H. CHAN, L. SHEN, Y. HSU, AND W. ISAAC TSENG, *An adaptive directional Haar framelet-
975 based reconstruction algorithm for parallel magnetic resonance imaging*, SIAM Journal on Imaging
976 Sciences, 9 (2016), pp. 794–821.
977 [30] D. LIANG, H. WANG, AND L. YING, *SENSE reconstruction with nonlocal TV regularization*, in Proceed-
978 ings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society
979 (IEEE EMBC), 2009, pp. 1032–1035.
980 [31] B. LIU, K. KING, M. STECKNER, J. XIE, J. SHENG, AND L. YING, *Regularized sensitivity encoding (sense)
981 reconstruction using bregman iterations*, Magnetic Resonance in Medicine, 61 (2009), pp. 145–152.
982 [32] G. LUO, X. WANG, V. ROELOFFS, Z. TAN, AND M. UECKER, *Joint estimation of coil sensitivities
983 and image content using a deep image prior*, in Proceedings of International Society for Magnetic
984 Resonance in Medicine (ISMRM), 2021.
985 [33] M. LUSTIG, D. DONOHO, AND J. M. PAULY, *Sparse MRI: The application of compressed sensing for
986 rapid MR imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195.
987 [34] M. LUSTIG, D. L. DONOHO, J. M. SANTOS, AND J. M. PAULY, *Compressed sensing MRI*, IEEE Signal
988 Processing Magazine, 25 (2008), pp. 72–82.
989 [35] M. LUSTIG AND J. M. PAULY, *SPIRiT: Iterative self-consistent parallel imaging reconstruction from
990 arbitrary k -space*, Magnetic Resonance in Medicine, 64 (2010), pp. 457–471.
991 [36] A. MAJUMDAR AND R. K. WARD, *Calibration-less multi-coil MR image reconstruction*, Magnetic Reso-
992 nance Imaging, 30 (2012), pp. 1032–1045.
993 [37] N. MENG, Y. YANG, Z. XU, AND J. SUN, *A prior learning network for joint image and sensitivity
994 estimation in parallel MR imaging*, in Proceedings of the International Conference on Medical Image
995 Computing and Computer-Assisted Intervention (MICCAI), 2019.
996 [38] P. NAIR, R. G. GAVASKAR, AND K. N. CHAUDHURY, *Fixed-point and objective convergence of plug-and-
997 play algorithms*, IEEE Transactions on Computational Imaging, 7 (2021), pp. 337–348.
998 [39] G. OH, B. SIM, H. CHUNG, L. SUNWOO, AND J. C. YE, *Unpaired deep learning for accelerated mri
999 using optimal transport driven cyclegan*, IEEE Transactions on Computational Imaging, 6 (2020),
1000 pp. 1285–1296.
1001 [40] J.-C. PESQUET, A. REPETTI, M. TERRIS, AND Y. WIAUX, *Learning maximally monotone operators for
1002 image recovery*, SIAM Journal on Imaging Sciences, 14 (2021), pp. 1206–1237.
1003 [41] N. G. POLSON, J. G. SCOTT, B. T. WILLARD, ET AL., *Proximal algorithms in statistics and machine*

- learning, Statistical Science, 30 (2015), pp. 559–581.
- [42] A. PRAMANIK, H. K. AGGARWAL, AND M. JACOB, *Deep generalization of structured low-rank algorithms (Deep-SLR)*, IEEE Transactions on Medical Imaging, 39 (2020), pp. 4186–4197.
- [43] K. P. PRUESSMANN, M. WEIGER, M. B. SCHEIDECKER, AND P. BOESIGER, *SENSE: Sensitivity encoding for fast MRI*, Magnetic Resonance in Medicine, 42 (1999), pp. 952–962.
- [44] Y. ROMANO, M. ELAD, AND P. MILANFAR, *The little engine that could: Regularization by denoising (RED)*, SIAM Journal on Imaging Sciences, 10 (2017), pp. 1804–1844.
- [45] E. RYU, J. LIU, S. WANG, X. CHEN, Z. WANG, AND W. YIN, *Plug-and-play methods provably converge with properly trained denoisers*, in Proceedings of the International Conference on Machine Learning (ICML), 2019, pp. 5546–5557.
- [46] J. SCHLEMPER, J. CABALLERO, J. V. HAJNAL, A. N. PRICE, AND D. RUECKERT, *A deep cascade of convolutional neural networks for dynamic MR image reconstruction*, IEEE Transactions on Medical Imaging, 37 (2017), pp. 491–503.
- [47] U. SCHMIDT AND S. ROTH, *Shrinkage fields for effective image restoration*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2774–2781.
- [48] H. SHE, R. CHEN, D. LIANG, E. V. DiBELLA, AND L. YING, *Sparse BLIP: Blind iterative parallel imaging reconstruction using compressed sensing*, Magnetic Resonance in Medicine, 71 (2014), pp. 645–660.
- [49] K. Y. SHENG FANG AND K. Y. SHENG FANG, *Coherence regularization for SENSE reconstruction with a nonlocal operator CORNOL*, Magnetic Resonance in Medicine, 64 (2010), pp. 1413–1425.
- [50] A. SRIRAM, J. ZBONTAR, T. MURRELL, A. DEFazio, C. L. ZITNICK, N. YAKUBOVA, F. KNOLL, AND P. JOHNSON, *End-to-end variational networks for accelerated MRI reconstruction*, in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2020.
- [51] J. SUN AND M. F. TAPPEN, *Learning non-local range Markov random field for image restoration*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 2745–2752.
- [52] 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, 7 (2021), pp. 849–863.
- [53] A. B. TAYLOR, J. M. HENDRICKX, AND F. GLINEUR, *Exact Worst-Case Convergence Rates of the Proximal Gradient Method for Composite Convex Minimization*, Journal of Optimization Theory and Applications, 178 (2018), pp. 455–476.
- [54] J. D. TRZASKO AND A. MANDUCA, *Calibrationless parallel MRI using CLEAR*, in Proceedings of the 2011 Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2011.
- [55] M. UECKER, T. HOHAGE, K. T. BLOCK, AND J. FRAHM, *Image reconstruction by regularized nonlinear inversion—joint estimation of coil sensitivities and image content*, Magnetic Resonance in Medicine, 60 (2008), pp. 674–682.
- [56] 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, 71 (2014), pp. 990–1001.
- [57] S. WANG, S. TAN, Y. GAO, Q. LIU, L. YING, T. XIAO, Y. LIU, X. LIU, H. ZHENG, AND D. LIANG, *Learning joint-sparse codes for calibration-free parallel MR imaging*, IEEE Transactions on Medical Imaging, 37 (2017), pp. 251–261.
- [58] D. WELLER, *Reconstruction with dictionary learning for accelerated parallel magnetic resonance imaging*, in Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 2016, pp. 105–108.
- [59] D. T. WYMER, K. P. PATEL, W. F. BURKE III, AND V. K. BHATIA, *Phase-contrast mri: physics, techniques, and clinical applications*, Radiographics, 40 (2020), pp. 122–140.
- [60] X. XIE, J. WU, G. LIU, Z. ZHONG, AND Z. LIN, *Differentiable linearized ADMM*, in Proceedings of the International Conference on Machine Learning (ICML), 2019, pp. 6902–6911.
- [61] B. XIN, Y. WANG, W. GAO, D. WIPF, AND B. WANG, *Maximal sparsity with deep networks?*, in Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS), 2016.
- [62] 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, 27 (2020), pp. 1280–1284.

- 1058 [63] Z. XU AND J. SUN, *Model-driven deep-learning*, National Science Review, 5 (2017), pp. 22–24.
- 1059 [64] C. YANG, Y. GU, B. CHEN, H. MA, AND H. C. SO, *Learning proximal operator methods for noncon-*
1060 *vex sparse recovery with theoretical guarantee*, IEEE Transactions on Signal Processing, 68 (2020),
1061 pp. 5244–5259.
- 1062 [65] Y. YANG, J. SUN, H. LI, AND Z. XU, *Deep ADMM-Net for compressive sensing MRI*, in Proceedings of
1063 the International Conference on Neural Information Processing Systems (NeurIPS), 2016, p. 10–18.
- 1064 [66] Y. YANG, J. SUN, H. LI, AND Z. XU, *ADMM-CSNet: A deep learning approach for image compressive*
1065 *sensing*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42 (2020), pp. 521–538.
- 1066 [67] L. YING AND J. SHENG, *Joint image reconstruction and sensitivity estimation in sense (JSENSE)*, Mag-
1067 *netic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance*
1068 *in Medicine*, 57 (2007), pp. 1196–1202.
- 1069 [68] J. ZHANG AND B. GHANEM, *ISTA-Net: Interpretable optimization-inspired deep network for image com-*
1070 *pressive sensing*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
(CVPR), 2018.
- 1071 [69] J. ZHANG, C. LIU, AND M. E. MOSELEY, *Parallel reconstruction using null operations*, Magnetic Reso-
1072 *nance in Medicine*, 66 (2011), pp. 1241–1253.
- 1073 [70] K. ZHANG, Y. LI, W. ZUO, L. ZHANG, L. VAN GOOL, AND R. TIMOFTE, *Plug-and-Play image restoration*
1074 *with deep denoiser prior*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2021),
1075 pp. 1–1.
- 1076 [71] P. ZHANG, F. WANG, W. XU, AND Y. LI, *Multi-channel generative adversarial network for parallel*
1077 *magnetic resonance image reconstruction in k-space*, in Proceedings of the International Conference
1078 *on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018, pp. 180–188.
- 1079