



Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2023 April ; 2023: . doi:10.1109/isbi53787.2023.10230471.

ACCELERATED PARALLEL MRI USING MEMORY EFFICIENT AND ROBUST MONOTONE OPERATOR LEARNING (MOL)

Aniket Pramanik,

Mathews Jacob

The University of Iowa, Iowa City, USA

Abstract

Model-based deep learning methods that combine imaging physics with learned regularization priors have been emerging as powerful tools for parallel MRI acceleration. The main focus of this paper is to determine the utility of the monotone operator learning (MOL) framework in the parallel MRI setting. The MOL algorithm alternates between a gradient descent step using a monotone convolutional neural network (CNN) and a conjugate gradient algorithm to encourage data consistency. The benefits of this approach include similar guarantees as compressive sensing algorithms including uniqueness, convergence, and stability, while being significantly more memory efficient than unrolled methods. We validate the proposed scheme by comparing it with different unrolled algorithms in the context of accelerated parallel MRI for static and dynamic settings.

1. INTRODUCTION

Compressive sensing (CS) algorithms have revolutionized accelerated Magnetic Resonance Imaging (MRI). These methods consider the recovery of images from noisy and highly undersampled multi-channel measurements by posing the recovery as a convex optimization problem. The benefits of CS algorithms include theoretical guarantees on uniqueness, convergence, and stability. In recent years, model-based deep-learning (MoDL) has gained tremendous success in parallel MRI [1, 2]. These algorithms alternate between a gradient descent step to minimize data consistency, followed by a denoising step using a CNN. The alternating optimization blocks are unrolled for few iterations and are trained end-to-end with weights shared across iterations. The image quality of the reconstructions offered by deep unrolled (DU) approaches often surpasses the ones from CS algorithms, in addition to being significantly more efficient in computation during inference.

The central challenge with DUs, which involves algorithm unrolling, is its high memory demand during training. This drawback restricts its utility in high-dimensional applications including 2D+time and 3D applications. Several strategies that trade computational complexity for reduce memory footprint have been introduced to reduce the memory

COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted using human subject data. Approval was granted by the Ethics Committee of the institute where the data were acquired.

demands [3, 4]. The deep equilibrium (DEQ) framework was introduced as a memory-efficient alternative where non-expansive optimization blocks are iterated until convergence to a fixed point [5]. Since DEQ relies on fixed point iterations rather than back-propagation, it reduces memory demand by a factor of the number of iterations, while its computational complexity is comparable to an unrolled network [6]; the tradeoffs offered by this scheme are better than the above discussed strategies [3, 4]. Unlike unrolled methods, the backpropagation steps in DEQ will be accurate only if the forward iterations converge. The convergence of several DEQ schemes were rigorously analyzed in [5, 7]. However, these theoretical convergence guarantees are unfortunately not valid in the highly undersampled setting, when the forward model is low-rank or ill-conditioned.

We had recently introduced a DEQ scheme to address the limitations of DUs in [8]. This algorithm inherits many of the desirable properties of CS methods including uniqueness of the solution, guaranteed convergence, and robustness to input perturbations. The approach relies on a forward-backward splitting algorithm, where the score function rather than its proximal map, is modeled by a CNN; the algorithm alternates between a gradient step to improve the prior probability and a conjugate gradient algorithm to enforce data consistency using a damping parameter α . We constrain the score network to be a monotone operator, which we show, is a necessary and sufficient condition for the fixed point of the iterations to be unique. Current algorithms such as MoDL and RED are special cases of the proposed algorithm, when the damping parameter $\alpha = 1$. Since the monotone operator is central to our approach, we call it as Monotone Operator Learning (MOL) [8]. We have introduced theoretical guarantees on the uniqueness of the fixed point of the algorithm, the convergence of the algorithm to the fixed point even when the forward model is low-rank, and the stability of the algorithm to adversarial perturbations. It is also shown that a monotone operator can be constructed as a residual CNN, where a Lipschitz constraint is enforced on the CNN. Traditional DEQ schemes use spectral normalization layer by layer which is a very conservative bound on Lipschitz of the CNN [5]; our experiments show that the use of this bound translates to poor reconstructions. We had introduced a Lipschitz regularized training loss (MOL-LR), which can offer a more realistic bound. While the basic theory was introduced in [8], the results in [8] were stated without proofs. In addition, the preliminary experimental results in [8] were restricted by the small 2D training dataset and the limited number of algorithms we compared against.

The main focus of this work is to rigorously validate the theoretical results in [8] using experimental data in the context of parallel MRI. We also show the preliminary utility of the algorithm in a 2D+time cardiac cine MRI application, where it is challenging to use unrolled algorithms because of memory constraints. We review the theoretical results [8] and provide full proofs in the arXiv version [9]; they are not included in this conference version because of space constraints. In this paper, we compare the proposed MOL-LR [8] reconstructions against the DEQ scheme (DE-GRAD) in [5], MOL-SN, UNET and DU approaches MoDL [2], ADMM-Net [10] in the context of 2D parallel MRI. We test our hypothesis to conclude that the performance of the proposed DEQ scheme is comparable to that of the unrolled algorithms, while being significantly more memory efficient. We also evaluate the sensitivity of the algorithms to Gaussian and worst-case input perturbations.

This experiment is expected to reveal the benefit of the theoretical bound on robustness in [8].

2. MONOTONE OPERATOR LEARNING

We will now briefly review the MOL algorithm introduced in [8]. The results in [8] were stated without proofs. We now restate the main results for accessibility, while a careful reader can find the proofs in the full arXiv version [9]. Consider recovery of an image $x \in \mathbb{C}$ from its noisy under-sampled measurements b such that $b = \mathcal{A}(x) + n$ where \mathcal{A} is a linear operator and n is Gaussian noise. The maximum a posteriori (MAP) estimate of x solves for,

$$C(x) = \arg \min_x \frac{\lambda}{2} \|Ax - b\|_2^2 + \phi(x) \quad (1)$$

where $\phi(x) = -\log(p(x))$ with $p(x)$ representing the prior probability density, while $\lambda = \frac{1}{\sigma^2}$ accounts for measurement noise. The minima of (1) satisfies the relation:

$$\underbrace{\lambda A^H(Ax - b)}_{\mathcal{F}(x)} + \underbrace{\nabla_x \phi(x)}_{\mathcal{F}(x)} = 0. \quad (2)$$

Here, \mathcal{F} is the gradient of $\phi(x)$ and is often termed as the score function. As discussed earlier, one of the strengths of CS formulations is the uniqueness of the solutions. The uniqueness of the solutions that satisfy (2) were analyzed in [8] using the theory of monotone operators, defined as

Assumption: The operator $\mathcal{F}: \mathbb{C}^M \rightarrow \mathbb{C}^M$ is m -monotone if:

$$\Re(\langle x-y, \mathcal{F}(x) - \mathcal{F}(y) \rangle) \geq m \|x - y\|_2^2, \quad m > 0. \quad (3)$$

for all $x, y \in \mathbb{C}^M$. Here, $\Re(x)$ denotes the real part of x . The following result from [9] provides the necessary and sufficient condition for the solution of (2) to be unique.

Proposition 2.1 [9] *The fixed point of (2) is unique for a specific b , iff \mathcal{F} is m -monotone with $m > 0$.*

Proposition 2.2 [9] *$\mathcal{F}: \mathbb{C}^M \rightarrow \mathbb{C}^M$ is m -monotone if it can be represented as a residual CNN, $\mathcal{F} = \mathcal{J} - \mathcal{H}_\theta$, where $\mathcal{H}_\theta: \mathbb{C}^M \rightarrow \mathbb{C}^M$ has a Lipschitz constant, $L[\mathcal{H}_\theta] = (1 - m); 0 < m < 1$.*

2.1. Proposed MOL algorithm

With these results, we use a forward-backward splitting of (2) for $\alpha > 0$, to yield the iterative algorithm $x_{n+1} = (I + \alpha\mathcal{F})^{-1}(I - \alpha\mathcal{F})(x_n)$, which is expanded as

$$x_{n+1} = \underbrace{(\mathcal{F} + \alpha\lambda\mathbf{A}^H\mathbf{A})^{-1}((1 - \alpha)x_n + \alpha\mathcal{H}_\theta(x_n))}_{\mathcal{F}_{\text{MOL}}(\mathbf{x}_n)} + \underbrace{(\mathcal{F} + \alpha\lambda\mathbf{A}^H\mathbf{A})^{-1}(\alpha\lambda\mathbf{A}^H\mathbf{b})}_{\mathbf{z}} \quad (4)$$

Proposition 2.3 Consider the algorithm specified by (4), where \mathcal{F} is an m -monotone operator. Assume that (4) has a fixed point specified by $\mathbf{x}^*(\mathbf{b})$. Then, convergence is guaranteed for an arbitrary \mathbf{A} operator when

$$\alpha < \frac{2m}{(2 - m)^2} = \alpha_{\text{max}} \quad (5)$$

The algorithms [2, 1, 11] correspond to the special case of $\alpha = 1$. Setting $\alpha = 1$ in (5), we see that the algorithm will converge if $m \geq 3 - \sqrt{5} = 0.76$ or $L[\mathcal{H}_\theta] \leq 0.24$. The denoising ability of a network is dependent on its Lipschitz bound; a smaller $L[\mathcal{H}_\theta]$ bound translates to poor performance of the resulting MOL algorithm. The use of the damping factor $\alpha < 1$ allows us to use denoising networks \mathcal{H}_θ with larger Lipschitz bounds and hence improved denoising performance. For instance, if we choose $m = 0.1$; $L[\mathcal{H}_\theta] = 0.9$, from (5), the algorithm will converge if $\alpha < 0.055$.

The following result bounds the perturbation in the solutions in response to worst case measurement noise in \mathbf{b} .

Proposition 2.4 [8] Consider \mathbf{z}_1 and \mathbf{z}_2 to be measurements with $\delta = \mathbf{z}_2 - \mathbf{z}_1$ as the perturbation. Let the corresponding outputs of the MOL algorithm be $\mathbf{x}^*(\mathbf{z}_1)$ and $\mathbf{x}^*(\mathbf{z}_2)$, respectively, with $\Delta = \mathbf{x}^*(\mathbf{z}_2) - \mathbf{x}^*(\mathbf{z}_1)$ as the output perturbation,

$$\|\Delta\|_2 \leq \frac{\alpha\lambda}{1 - \sqrt{1 - 2\alpha m + \alpha^2(2 - m)^2}} \|\delta\|_2. \text{ When } \alpha \text{ is small, we have } \lim_{\alpha \rightarrow 0} \|\Delta\|_2 \leq \frac{\lambda}{m} \|\delta\|_2.$$

We note from Proposition 2.2 that a monotone $\mathcal{F} = \mathcal{F} - \mathcal{H}_\theta$ can be learned by constraining the Lipschitz constant of \mathcal{H}_θ . A common approach is spectral normalization [5, 7]. However, this is a conservative estimate, often translating to lower performance. We use the Lipschitz

estimate [12], $L[\mathcal{H}_\theta] = \max_{\mathbf{x} \in \mathcal{S}} \sup_{\eta} \frac{\overbrace{\|\mathcal{H}_\theta(\mathbf{x} + \eta) - \mathcal{H}_\theta(\mathbf{x})\|_2}^{P(\mathbf{x})}}{\|\eta\|_2^2}$. This estimate is less conservative than

the one using spectral normalization. Our experiments show that the use of this estimate can indeed result in algorithms with convergence and robustness as predicted by the theory.

In the supervised learning setting, we propose to minimize, $V(\theta) = \sum_{i=0}^{N_t} \|\mathbf{x}_i^* - \mathbf{x}_i\|_2^2$ s.t.

$P(\mathbf{x}_i^*) \leq T; i = 0, \dots, N_i$. Here, $T = 1 - m$. The above loss function is minimized with respect to parameters θ of the CNN \mathcal{H}_θ using a log-barrier approach,

$$\theta^* = \arg \min_{\theta} \sum_{i=0}^{N_i} \underbrace{(\|\mathbf{x}_i^* - \mathbf{x}_i\|_2^2 - \beta \log(T - P(\mathbf{x}_i^*)))}_{C_i}. \quad (6)$$

\mathbf{x}_i^* is a fixed point of (4) that is dependent on the CNN parameters θ . $\mathbf{x}_i; i = 0, \dots, N_i$ and $\mathbf{b}_i; i = 0, \dots, N_i$ are the ground truth images in the training dataset and their undersampled measurements, respectively.

3. EXPERIMENTS AND RESULTS

The experiments in this paper build upon the preliminary results in [8], where the proposed MOL-LR (Lipschitz regularized) algorithm was compared against MOL-SN (spectral normalization) and MoDL in terms of performance. However, the models were trained with few datasets (90 slices from each of the 4 training subjects). In addition, the experiments in [8] did not study the robustness of the proposed algorithm shown in Proposition 2.4 or the benefit of the scheme in high-dimensional applications, in comparison to other unrolled methods. In this work, we address above limitations of [8] by performing experiments on a large publicly available brain MRI data from the Calgary Campinas Dataset (CCP) [13]. CCP consists of 12-channel (coil) T1-weighted 2D brain data from 117 subjects. We split data from 40, 7 and 20 subjects into training, validation and testing sets respectively. Cartesian 2D non-uniform variable density masks are used for 4-fold undersampling of the datasets.

We compare the performance of MOL-LR and its robustness to against DEQ methods (DE-GRAD, MOL-SN), UNET and DUs (MoDL, ADMM-Net) in the first row of images in Fig. 1 for recovery from 4-fold undersampled data. The SENSE, MOL-SN and DE-GRAD reconstructions are relatively more noisy compared to other methods. SENSE is a CS algorithm, utilizing coil sensitivities obtained from calibration scans. It fails to perform at such a high acceleration factor. MOL-SN and DE-GRAD are DEQs with Lipschitz of the network bounded by spectral normalization of each layer. Since, this is a very strict bound, the CNNs attain a much lower Lipschitz compared to MOL-LR, which translates to poor performance. MOL-LR performs at par with DUs (10-iterations of MoDL, ADMM-Net) and outperforms UNET.

We study degradation of the reconstructions, when the input measurements are corrupted by Gaussian or worst-case perturbations determined using an adversarial attack. The experiments in the third row of Fig. 1 (a) show that all the methods perform similarly on Gaussian noise. By contrast, the experiments in the second row of Fig. 1 (a) show that MOL-SN and MOL-LR are relatively more robust to adversarial noise, compared to MoDL, UNET, and ADMM-Net. Both MOL-SN and MOL-LR are associated with guarantees on

robustness. The plots in Fig. 1 (b) shows drop in PSNR with respect to percentage of Adversarial/Gaussian noise (0% to 20% for $\epsilon = [0.0, 0.2]$) and shows similar trend.

We demonstrate the benefit of reduced memory demand in MOL over unrolled networks through experiments on 2D+time cardiac data, run on a 16GB GPU. While MOL-LR 3D performs recovery in 3D, MoDL [2] could do it only in 2D due to lack of memory for multiple unrolls. MOL-LR 3D outperforms MoDL 2D as shown in Fig. 2. MoDL 2D's poor performance can be attributed to lack of temporal information which is well utilized by MOL-LR 3D.

4. CONCLUSION

In this paper, we demonstrate the benefits of a DEQ-based monotone operator learning (MOL), proposed in [8]. Similar to compressed sensing algorithms with convex priors, MOL possesses guarantees for uniqueness of the solution, convergence to the fixed-point and stability to input perturbations. MOL is found to have ten-fold reduction in memory consumption compared to one consumed by 10-iterations of unrolled networks. Results show that MOL is significantly more robust to adversarial attacks and performs at par on noiseless data, compared to unrolled networks. We apply MOL to higher-dimensional problems which cannot be done for unrolled networks due to memory constraints.

Acknowledgments

This work is supported by NIH R01AG067078.

REFERENCES

- [1]. Hammernik et al. , "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018. [PubMed: 29115689]
- [2]. Aggarwal et al. , "MoDL: Model-based deep learning architecture for inverse problems," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 394–405, 2018. [PubMed: 30106719]
- [3]. Chen et al. , "Training deep nets with sublinear memory cost," *arXiv preprint arXiv:1604.06174*, 2016.
- [4]. Kellman et al. , "Memory-efficient learning for large-scale computational imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1403–1414, 2020.
- [5]. Gilton et al. , "Deep equilibrium architectures for inverse problems in imaging," *arXiv preprint arXiv:2102.07944*, 2021.
- [6]. Bai et al. , "Deep equilibrium models," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [7]. Ryu et al. , "Plug-and-play methods provably converge with properly trained denoisers," in *International Conference on Machine Learning*. PMLR, 2019, pp. 55465557.
- [8]. Pramanik Aniket and Jacob Mathews, "Improved Model Based Deep Learning Using Monotone Operator Learning (Mol)," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–4.
- [9]. Pramanik Aniket and Jacob Mathews, "Stable and memory-efficient image recovery using monotone operator learning (MOL)," *arXiv preprint arXiv:2206.04797*, 2022.
- [10]. Sun et al. , "Deep ADMM-Net for compressive sensing MRI," *Advances in neural information processing systems*, vol. 29, 2016.
- [11]. Romano et al. , "The little engine that could: Regularization by denoising (RED)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

- [12]. Bungert et al. , “CLIP: Cheap Lipschitz training of neural networks,” in International Conference on Scale Space and Variational Methods in Computer Vision. Springer, 2021, pp. 307–319.
- [13]. Souza at al., “An open, multi-vendor, multi-fieldstrength brain MR dataset and analysis of publicly available skull stripping methods agreement,” *NeuroImage*, vol. 170, pp. 482–494, 2018.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

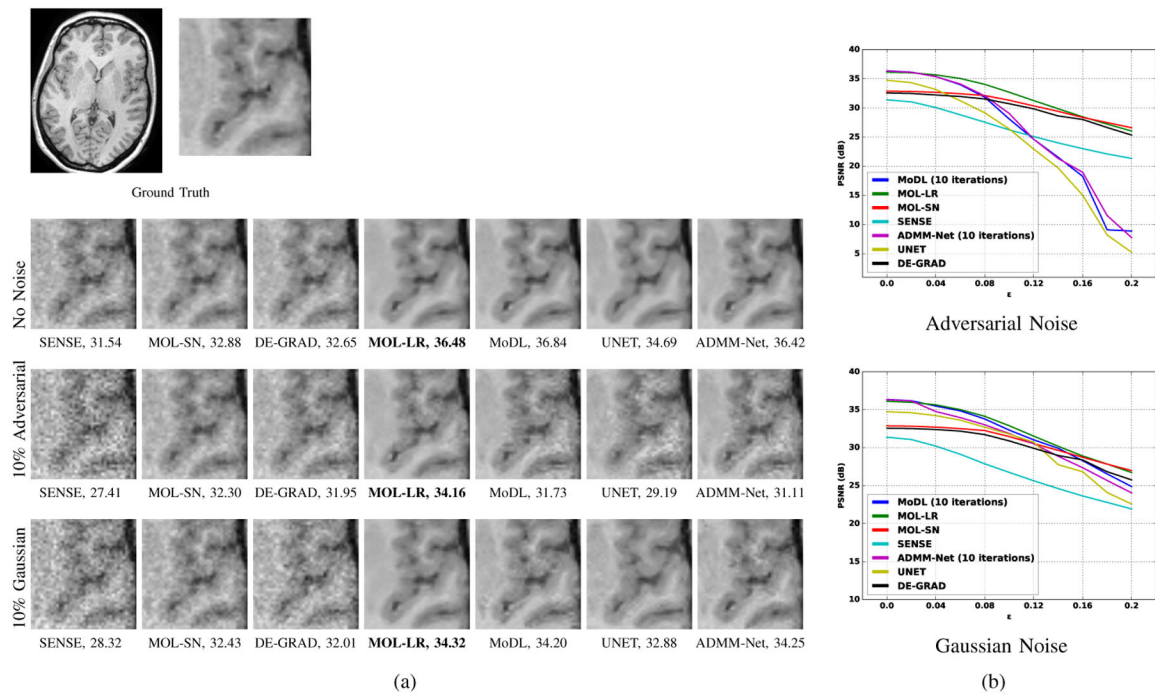


Fig. 1. Sensitivity of the algorithms to input perturbations. In (a), the rows correspond to reconstructed magnitude results of 4x accelerated multi-channel brain data for no noise, 10% worst-case (adversarial) and 10% gaussian noise, respectively. The PSNR (dB) values are reported for each case. The data was undersampled using a Cartesian 2D non-uniform variable-density mask; (b) shows plots of PSNR vs percentage of Adversarial or Gaussian Noise in terms of ϵ (from 0% to 20%).

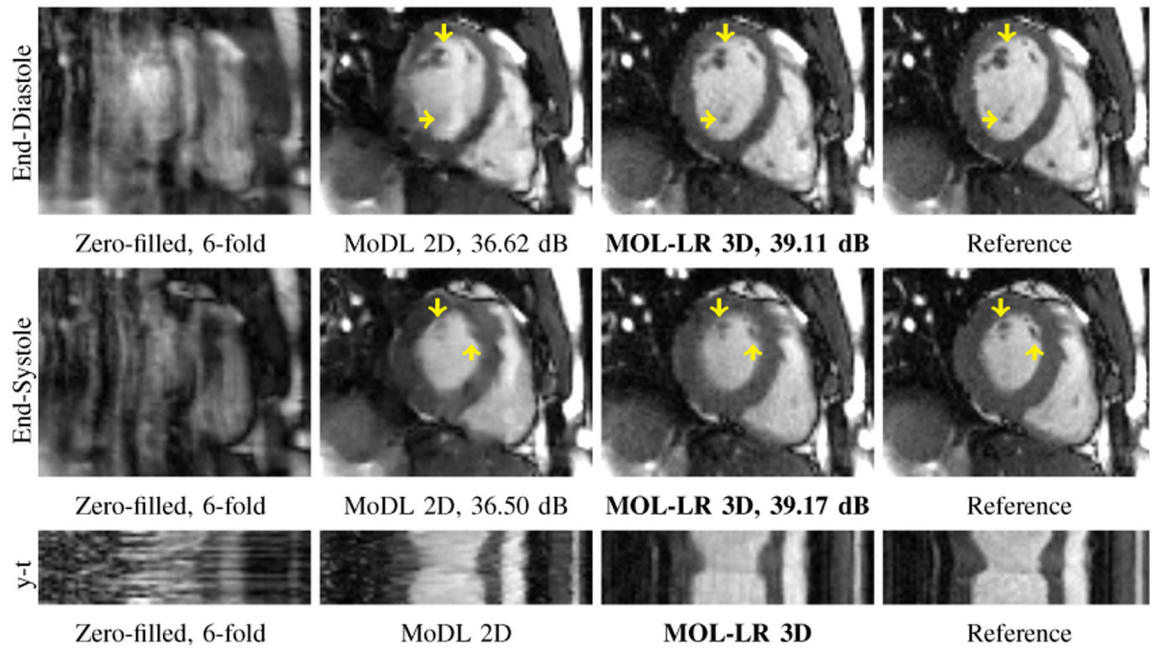


Fig. 2. MOL recovery of 2D+time cine data at 6x acceleration. PSNR (dB) values are reported for each case. The data is retrospectively undersampled using a Poisson density sampling pattern.