IMPROVED MODEL BASED DEEP LEARNING USING MONOTONE OPERATOR LEARNING (MOL)

Aniket Pramanik, Mathews Jacob

The University of Iowa, Iowa City, USA

ABSTRACT

Model-based deep learning (MoDL) algorithms that rely on unrolling are emerging as powerful tools for image recovery. In this work, we introduce a novel monotone operator learning framework to overcome some of the challenges associated with current unrolled frameworks, including high memory cost, lack of guarantees on robustness to perturbations, and low interpretability. Unlike current unrolled architectures that use finite number of iterations, we use the deep equilibrium (DEQ) framework to iterate the algorithm to convergence and to evaluate the gradient of the convolutional neural network blocks using Jacobian iterations. This approach significantly reduces the memory demand, facilitating the extension of MoDL algorithms to high dimensional problems. We constrain the CNN to be a monotone operator, which allows us to introduce algorithms with guaranteed convergence properties and robustness guarantees. We demonstrate the utility of the proposed scheme in the context of parallel MRI.

1. INTRODUCTION

The recovery of biomedical images from few measurements using compressive sensing (CS) algorithm has received considerable attention in the last decade, with wide ranging applications including MRI and microscopy. In the recent years, deep learning methods have been revolutionizing image recovery. Model based (MoDL) architectures, which unroll an iterative CS algorithm and replace the proximal mapping blocks by convolutional neural network (CNN) modules have been introduced by several authors including [1, 2]. While these schemes offer good performance and a significant reduction in training data demand, they have some limitations that restrict their utility. The main challenge in the training of these networks is the high memory demand. In particular, the images and features at each unrolling step need to be stored during back-propagation, which severely limits the number of unrolling steps in multi-dimensional applications. Another challenge is the interpretability of these models. Eventhough they are motivated by iterative algorithms, the limited number of unrolling steps often cannot guarantee the convergence of the algorithms to the minimum of any specific cost function. It is desirable to learn MoDL frameworks that inherit the desirable properties of classical convex optimization schemes.

Building upon deep equilibrium (DEQ) schemes that iterate a single layer [3] to convergence, a DEQ architecture was recently introduced for image recovery [4]. Assuming the network blocks to be non-expansive, the forward iterations are iterated until convergence [4]. Instead of relying on unrolling for backpropagation, DEQ schemes rely on Jacobian backward iterations to compute the gradient of the CNN block. The main benefit of this approach is the quite significant reduction in the memory demand. Because the Lipshitz constant of the CNN block should be less than one for the forward and backward iterations to converge, spectral normalization was used in [4]. We note that the Lipshitz constant computed using spectral methods of each layer is a very conservative upper bound. Our experiments in the the parallel MRI context show that the use of spectral normalization results in networks with limited expressive power, which translates to poor performance. Another challenge is that many algorithms such as ADMM and proximal gradients do not come with convergence guarantees in the MoDL setting when the data consistency (DC) term is not strictly convex (i.e. rank deficient forward model) [4, 5].

In this work, we introduce a novel DEQ framework for inverse problems, where we focus on the learning of a deep monotone operator [6]; we term this approach as monotone operator learning (MOL). We note that the derivative of convex priors are monotone operators. Even though the converse is not true, monotone operators can offer several benefits in the DEQ context, including guaranteed convergence and robustness to input perturbations. We introduce a forward backward algorithm, which uses the monotone operator. The proposed forward backward algorithm comes with convergence guarantees, provided the operator is strictly monotone. We show that any general residual network of the form \((I - H)\) involving a deep-network \(H\) is strictly monotone, provided the Lipshitz constant of \(H\) is less than unity. Because spectral normalization severely restricts the representation power, we propose an additional Lipshitz regularization term to the training loss. Our experiments show that this approach can offer significantly improved results over spectral normalized networks.

We demonstrate the utility of the proposed scheme in the context of parallel MRI. Our results show that the proposed scheme can offer results that are comparable to unrolled net-
works, albeit with quite significant reduction in memory demand. This work paves the way for the extension of model based deep learning schemes to multi-dimensional imaging (e.g. 3D or 3D+time) applications.

2. METHODS

2.1. Review of Deep Equilibrium Models

We consider the recovery of an image \( f : \mathbb{R}^2 \rightarrow \mathbb{C} \) from its undersampled linear measurements \( b = A(f) + n \), where \( A \) is a rank-deficient linear operator with maximum eigen value of 1 and \( n \) is noise. CS methods pose the recovery as an optimization problem with a convex prior

\[
    f^* = \arg \min_f \frac{\lambda}{2} \| A(f) - b \|^2 + \varphi(f) \quad (1)
\]

Since \( \varphi \) is often non-differentiable, iterative algorithms that rely on the proximal mapping of \( \varphi \) were introduced.

In the recent years, several authors have introduced model-based deep learning (MoDL) algorithms, where an iterative compressed sensing algorithm is unrolled and the proximal operators are replaced with off-the-shelf denoisers or learned CNN modules. While these strategies are powerful, they suffer from high memory demand and low interpretability. The deep equilibrium algorithm was introduced to reduce the memory demand of MoDL algorithms. When \( \varphi(f) \) is differentiable, the minimization of (1) satisfies the equation

\[
    \nabla f C = \lambda A^T (Af - b) + \partial \varphi(f) = 0 \quad (2)
\]

Gilton et al. [4], used iterative gradient descent \( f_{n+1} = f_n - \gamma \nabla C(f_n) \) algorithm, accelerated with Andersen iterations, to obtain the fixed point of the above equation. When \( A \) is a rank deficient operator, the forward and backward iterations are guaranteed to converge when the Lipshitz constant of the network \( H = (F - I) \) is less than unity and the step size \( \gamma < 1/2 \). With this setting, DEQ schemes compute the gradient of \( F \) using Jacobian iterations. This approach eliminates the need for algorithm unrolling, thus quite significantly reducing the memory demand of the algorithm; a single data consistency block and deep learning block \( F \) is sufficient to implement the algorithm. While other algorithms such as ADMM and proximal gradients can be used [4][5], the convergence guarantees are not valid when \( A \) is rank deficient.

2.2. Monotone operator learning and proposed algorithm

To overcome the above challenges, we focus on choosing \( F \) in (2) as a monotone operator [6]. An operator is \( m \)-monotone if it satisfies the condition:

\[
    \langle (x - y), F(x) - F(y) \rangle > m \| x - y \|^2 \quad (3)
\]

We note that the differential \( F = \partial \varphi \) is monotone if \( \varphi \) is convex. Although the converse is not true in general, DEQ schemes with monotone operator can guarantee convergence

\[
    f_{k+1} = (I + \alpha G)^{-1} \left( f_k - \alpha F(f_k) \right) \quad (4)
\]

with the following convergence guarantee:

Proposition 2.1 [5]. The forward backward optimization algorithm will converge to the fixed point of (3) if \( \alpha < 2m/L^2 \), where \( F \) is \( m \)-monotone with a Lipshitz constant of \( L \).

2.3. Residual network is a monotone operator

We note that [6] relies on a specially designed network layer that guarantee that \( F \) is monotone. Unfortunately, such single layer networks have low representation power to be used in our setting. We instead show that a residual network \( F = I - H \) is \( m \)-monotone, under the following condition.

Proposition 2.2 \( F \) is an \( m \)-monotone network if the Lipshitz constant of \( H \) is less than \( 1 - m \). The Lipshitz constant of \( F \) is bounded by \( (2 - m) \).

We hence substitute for \( F \) in (4) to obtain

\[
    x_{k+1} = Q^{-1} \left( 1 - \alpha \right) x_k + \alpha H(x_k) + \alpha \lambda A^H b \quad (5)
\]

where \( Q = (I + \alpha \lambda A^H A) \). We thus obtain the result:

Proposition 2.3 The iterative algorithm specified by (5) will converge to the fixed point of (3), denoted by \( f^*(b) \) if \( \alpha < 2m/(2 - m)^2 \).

We note that when \( \alpha = 1 \), the algorithm simplifies to the iterative approach in [2]; this algorithm will converge in a DEQ setting when \( m > 0.76 \) or when the Lipshitz constant of \( H \) is less than 0.24.

2.4. Robustness to input perturbations

A challenge with deep learning networks is their vulnerability to adversarial perturbations. The use of monotone operators allows us to bound the norm of the perturbations in the output in terms of the perturbations in the measurements.
Proposition 2.4 Let \( f^*(a) \) and \( f^*(b) \) be the fixed points corresponding to two set of measurements, denoted by \( a \) and \( b \), respectively. We then have
\[
\| f^*(a) - f^*(b) \|_2 \leq \frac{\alpha \lambda}{1 - \sqrt{1 - 2\alpha m + \alpha^2(2m)^2}} \| a - b \|_2
\]
(6)

When \( \alpha \to 0 \), we have \( \| f^*(a) - f^*(b) \|_2 \leq \lambda \| a - b \|_2/m \).

This result shows that \( m \) plays a very important role in the robustness of the system. In particular, higher Lipshitz constant of \( H \) implies networks with improved representation power, and hence can potentially offer higher performance. However, the improved performance will come with a smaller \( m \) and hence higher sensitivity to input perturbations.

2.5. Local Lipshitz regularization

The popular approach for constraining the Lipshitz bound of a network is spectral normalization of each layer. However, the product of the spectral bounds results in a very conservative estimate for the Lipshitz constant of the network; using spectral normalization results in networks with poor representation power, and hence worse results than unrolled networks that do not require these constraints.

In this work, we propose to use a regularization term to constrain the Lipshitz constant of \( H \). Instead of global Lipshitz constants over the entire domain of \( H \), we focus on evaluating the local Lipshitz constants to the subregion of the domain where the images live. Specifically, for each batch, we solve an optimization problem involving an adversarial perturbation of the fixed point:
\[
l(f^*) = \max_{\epsilon} \frac{\| H(f^* + \epsilon) - H(f) \|_2^2}{\| \epsilon \|_2^2}
\]
(7)

We optimize the above expression using a few gradient ascent steps during each epoch. The total loss at each epoch is the sum of the MSE loss obtained by comparing the reconstructions to the true images and the estimated Lipshitz constants of \( H \).
\[
\mathcal{L} = \| f^*(A^H b) - f_{\text{orig}} \|^2_2 + l(f^*)
\]
(8)

3. EXPERIMENTS AND RESULTS

The experiments were performed on 3D CUBE brain MRI dataset collected from 9 subjects at the University of Iowa using a 12-channel head coil. The dataset was split into 4 subjects for training, 2 for validation and the remaining 3 for testing. We chose the middle slices from each subject. 2D non-uniform variable density undersampling masks with different acceleration factors were used for the experiments. We used a five layer deep network as in [2] to implement \( H \). The parameter \( \alpha \) was chosen as \( 2m/(2-m)^2 \) to guarantee convergence. We did not use batch normalization because it is difficult to restrict the Lipshitz constant of \( H \).

We characterize the MOL framework in Fig. 2. It is observed from (a)-(e) that the MOL approach without spectral normalization or Lipshitz regularization (red curve) diverges rapidly. In particular, the forward iterations (nFE) increase to the maximum without converging, when back-propagation using Jacobian iterations provide erroneous gradients. It is observed that the forward iterations in the MOL approach with spectral normalization of \( H \) (MOL-SN) converges rapidly, as seen from the low number of forward iterations (nFE). However, this approach results in higher training and testing errors than the unrolled MoDL. This can be attributed to the low local Lipshitz constants (hence low representation power) of the \( H \) network. By contrast, the Lipshitz regularization (MOL-LR) allows us to keep the Lipshitz constant around 0.9, which offers higher representation power and image quality comparable to MoDL with 2 iterations.

| Methods  | 4x PSNR (dB) | 4x SSIM | 6x PSNR (dB) | 6x SSIM |
|----------|--------------|---------|--------------|---------|
| SENSE    | 37.39        | 0.983   | 34.04        | 0.967   |
| MoDL     | 41.02        | 0.992   | 37.98        | 0.985   |
| MOL-SN   | 38.31        | 0.985   | 35.19        | 0.972   |
| MOL-LR   | 40.75        | 0.991   | 37.71        | 0.984   |

Table 1. Quantitative comparison of SENSE, MoDL, MOL-SN and MOL-LR reconstructions. Average Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are reported for 30 test slices.
Fig. 3. Reconstruction results of 6-fold accelerated brain data. Top row are zoomed sections of the reconstructions. The proposed monotone operator learning with Lipshitz regularisation (MOL-LR) is compared against SENSE, unrolled MoDL with 10 iterations and MOL-SN reconstructions. MOL-LR is a Lipschitz regularized network to ensure convergence while MOL-SN performs spectral normalization of the network weights. The top row displays magnitude of coil sensitivity weighted images and the bottom row shows corresponding error images. The PSNR in dB is reported for the slice in each case. Proposed MOL-LR is highlighted in bold. The GPU memory demand of MOL-LR is ten times lower than that of MODL with 10 unrolling steps, while its performance is comparable.

10 iteration MoDL scheme does not require any constraints on the networks and hence can offer marginally higher performance. However, its Lipshitz constant is higher than unity and hence we dont have any guarantees on robustness as in [6]. In addition, the memory footprint of MOL is ten times smaller.

Comparison of the methods are shown in Fig. 3 and Table 1 respectively. Results show that the proposed DEQ scheme is able to offer similar performance compared to MoDL with 10 iterations. The memory footprint of MOL-LR is around ten times smaller than MoDL, making it an attractive option in higher dimensional problems. The MoDL scheme cannot be directly applied in 3D/4D setting due to its high memory demand.

4. CONCLUSION

We introduced a monotone operator learning framework for improved model-based deep learning of inverse problems. The proposed framework comes with guaranteed convergence and robustness. We introduced a Lipshitz regularization scheme to control the monotonicity parameter, which offers a trade-off between performance and robustness. The application of the framework to parallel MRI reconstruction setting shows that the performance of the monotone operator learning framework can match the quality of unrolled approaches, albeit with around 10 fold reduction in memory requirement. This work will enable the extension of MoDL schemes to multi-dimensional problems.

5. 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.

6. REFERENCES

[1] Karol Gregor and Yann LeCun, “Learning fast approximations of sparse coding,” in Proceedings of the 27th international conference on international conference on machine learning, 2010, pp. 399–406.

[2] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob, “MoDL: Model-based deep learning architecture for inverse problems,” IEEE transactions on medical imaging, vol. 38, no. 2, pp. 394–405, 2018.

[3] Shaojie Bai, Zico Kolter, and Vladlen Koltun, “Deep Equilibrium Models,” Proceedings of Neurips, 2019.

[4] Davis Gilton, Gregory Ongie, and Rebecca Willett, “Deep equilibrium architectures for inverse problems in imaging,” arXiv preprint arXiv:2102.07944, 2021.

[5] Ryu et al., “Plug-and-play methods provably converge with properly trained denoisers,” in International Conference on Machine Learning. PMLR, 2019, pp. 5546–5557.
[6] Ezra Winston and J Zico Kolter, “Monotone operator equilibrium networks,” *arXiv preprint arXiv:2006.08591*, 2020.