Unsupervised Deep Basis Pursuit: Learning inverse problems without ground-truth data

Jonathan I. Tamir[1], Stella X. Yu[1,2], Michael Lustig[1]

***Abstract*—Basis pursuit is a compressed sensing optimization in which the ℓ_1-norm is minimized subject to model error constraints. Here we use a deep neural network prior instead of ℓ_1-regularization. Using known noise statistics, we jointly learn the prior and reconstruct images without access to ground-truth data. During training, we use alternating minimization across an unrolled iterative network and jointly solve for the neural network weights and training set image reconstructions. At inference, we fix the weights and pass the measurements through the network. We compare reconstruction performance between unsupervised and supervised (i.e. with ground-truth) methods. We hypothesize this technique could be used to learn reconstruction when ground-truth data are unavailable, such as in high-resolution dynamic MRI.**

I. Introduction

Deep learning in tandem with model-based iterative optimization [2]–[6], i.e. model-based deep learning, has shown great promise at solving imaging-based inverse problems beyond the capabilities of compressed sensing [7]. These networks typically require hundreds to thousands of examples for training, consisting of pairs of corrupted measurements and the desired ground-truth image. The reconstruction is then trained in an end-to-end fashion, in which data are reconstructed with the network and compared to the ground-truth result. in many cases, collecting a large set of fully sampled data for training is expensive, impractical, or impossible.

In this work, we present an approach to model-based deep learning without access to ground-truth data [8]–[10]. We take advantage of (known) noise statistics for each training example and formulate the problem as an extension of basis pursuit denoising [11] with a deep convolutional neural network (CNN) prior in place of image sparsity. During training, we jointly solve for the CNN weights and the reconstructed training set images. At inference time, we fix the weights and pass the measured data through the network.

As proof of principle, we apply the technique to under-sampled, multi-channel magnetic resonance imaging (MRI). We compare our Deep Basis Pursuit (DBP) formulation with and without supervised learning, as well as to MoDL [6], a recently proposed unrolled model-based network that uses ground-truth data for training. We show that in the unsupervised setting, we are able to approach the image reconstruction quality of supervised learning, thus opening the door to applications where collecting fully sampled data is not possible.

[1]Electrical Engineering and Computer Sciences, University of California, Berkeley. [2]International Computer Science Institute, University of California, Berkeley. Correspondence to: Jonathan Tamir <jtamir@eecs.berkeley.edu>. This work was first presented at the ISMRM 27th Annual Meeting, Montreal, Canada 2019 [1].

II. Background

A. Imaging Model

We focus on the discretized linear signal model under additive white Gaussian noise:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{C}^N$ is the vectorized unknown image, $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ is the discretized forward model describing the imaging system, $\boldsymbol{y} \in \mathbb{C}^M$ is a vector of the acquired measurements, and $\boldsymbol{v} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$ is a complex-valued Gaussian noise vector. We are interested in the ill-posed regime, where $M < N$. To make the inverse problem well-posed, $\boldsymbol{x}$ is commonly solved through a regularized least-squares:

$$\arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right\|_2^2 + \lambda Q(\boldsymbol{x}), \tag{2}$$

where $Q(\boldsymbol{x})$ is a suitable regularization term, and $\lambda > 0$ is the strength of the regularization.

An alternative, equivalent formulation that directly accounts for the model error due to noise is the constrained problem:

$$\begin{aligned} \arg\min_{\boldsymbol{x}} \quad & Q(\boldsymbol{x}) \\ \text{subject to} \quad & \left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right\|_2 \leq \epsilon, \end{aligned} \tag{3}$$

where $\epsilon = \sigma\sqrt{M}$ is the square-root of the expected noise power in the measurements. When an ℓ_1-norm is used for regularization, this is known as basis pursuit denoising [11], and provides an intuitive formulation as it finds the best (sparsist) representation given a noise error constraint.

B. Deep Learning Image Reconstruction

CNNs have recently been used to solve imaging inverse problems, relying on the network architecture and training data to learn the inverse mapping. When a large corpus of training data is available, it is possible to learn the inverse mapping directly from under-sampled measurements, typically by first transforming the measurements to the image domain either through the adjoint operation $\boldsymbol{A}^*\boldsymbol{y}$ or through a conventional reconstruction. Except for the initial transformation, these models do not take advantage of knowledge of the imaging system in the network

architecture. Thus, they require substantial training data and are prone to overfitting and CNN artifacts [12].

More recently, network architectures that combine both CNN blocks and data consistency blocks incorporating knowledge of the forward model have grown in popularity, as they allow for robustness against CNN artifacts and training with limited data [5], [6]. These architectures are inspired by conventional first-order iterative algorithms intended to solve the unconstrained problem (2), and typically alternate between data consistency and manifold projection. To facilitate training with backpropagation, the iterative algorithms are unrolled for a finite number of steps and optimized in an end-to-end manner. As the network is differentiable, gradient updates can be computed through the application of the forward operator with auto-differentiation.

For a particular network architecture, we can view the image reconstruction as a feed-forward network

$$\hat{\boldsymbol{x}}_{\boldsymbol{w}} = \mathcal{F}_{\boldsymbol{w}}\left(\boldsymbol{y}; \boldsymbol{A}\right), \tag{4}$$

where $\mathcal{F}_{\boldsymbol{w}}$ is a deep network parameterized by weights $\boldsymbol{w}$ that operates on the measurements and optionally incorporates knowledge of the forward model. Given a training set of inputs $\{\boldsymbol{y}^{(i)}, \boldsymbol{A}^{(i)}\}_{i=1}^{L}$ and corresponding ground-truth images $\{\boldsymbol{x}^{(i)}\}_{i=1}^{L}$, the network weights can be trained in a traditional end-to-end fashion by minimizing the average training loss as measured by the loss function $\mathcal{L}$:

$$\min_{\boldsymbol{w}} \frac{1}{L} \sum_{i=1}^{L} \mathcal{L}\left(\hat{\boldsymbol{x}}_{\boldsymbol{w}}^{(i)}, \boldsymbol{x}^{(i)}\right). \tag{5}$$

For inference, the weights are fixed and new measurements are reconstructed through a forward pass of (4).

III. Proposed Method

Inspired by other model-based deep learning architectures [2]–[6], we propose a new unrolled network based on basis pursuit denoising, which we call Deep Basis Pursuit (DBP). We assume the noise statistics of the measurements are known and we use them to self-regularize the solution. In turn, we propose to train in an unsupervised fashion in the measurement domain, taking advantage of explicit control of the error between the measurements and the output of the network. We first describe the DBP model, and then discuss training the in an unsupervised fashion without ground-truth data.

A. Deep Basis Pursuit

We combine the data consistency constraint of basis pursuit denoising (3) with the ℓ_2-norm incorporating a CNN auto-encoder. The DBP optimization is given by

$$\begin{aligned} \arg\min_{\boldsymbol{x}} \quad & \frac{1}{2}\left\|\mathcal{N}_{\boldsymbol{w}}\left(\boldsymbol{x}\right)\right\|_2^2 \\ \text{subject to} \quad & \left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right\|_2 \leq \epsilon, \end{aligned} \tag{6}$$

where $\mathcal{N}_{\boldsymbol{w}}(\boldsymbol{x}) \equiv \boldsymbol{x} - \mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x})$ is a CNN parameterized by weights $\boldsymbol{w}$ that aims to estimate noise and aliasing [4], [6]. In other words, $\mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x})$ represents a denoised version of $\boldsymbol{x}$. In this way, we seek to find the "cleanest" representation of $\boldsymbol{x}$ while allowing for the expected data inconsistency due to noise.

To approximately solve (6), we consider an alternating minimization [6], repeated N_1 times:

$$\boldsymbol{r}_k = \mathcal{R}_{\boldsymbol{w}}(\boldsymbol{x}_{k-1}), \tag{7}$$

$$\boldsymbol{x}_k = \arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{r}_k\right\|_2^2 \text{ s.t. } \left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right\|_2 \leq \epsilon. \tag{8}$$

Subproblem (7) is a forward pass through the CNN. Subproblem (8) is convex and can solved with ADMM [13]. We introduce the slack variable $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$ and the dual variable $\boldsymbol{u}$, and apply the following update steps, repeated N_2 times:

$$\boldsymbol{x}_l = \left(\rho\boldsymbol{A}^*\boldsymbol{A} + \boldsymbol{I}\right)^{-1}\left(\rho\boldsymbol{A}^*\left(\boldsymbol{z}_{l-1} - \boldsymbol{u}_{l-1}\right) + \boldsymbol{r}_k\right), \tag{9}$$

$$\boldsymbol{z}_l = \boldsymbol{y} + \mathrm{L2Proj}\left(\boldsymbol{A}\boldsymbol{x}_l + \boldsymbol{u}_{l-1} - \boldsymbol{y}, \epsilon\right), \tag{10}$$

$$\boldsymbol{u}_l = \boldsymbol{u}_{l-1} + \boldsymbol{A}\boldsymbol{x}_l + \boldsymbol{z}_l, \tag{11}$$

where $\rho > 0$ is the ADMM penalty parameter and $\mathrm{L2Proj}(\boldsymbol{z}, \epsilon)$ is the projection of $\boldsymbol{z}$ onto the ℓ_2-ball of radius ϵ. The update steps are amenable to matrix-free optimization, as the forward and adjoint calculations can be represented as computationally efficient operators. In particular, subproblem (9) can be approximately solved with N_3 iterations of the Conjugate Gradient Method.

Altogether, we can view DBP as an unrolled optimization alternating between CNN layers and data consistency layers, as shown in Fig. 1. At each layer, the same CNN is used, though it is possible in general to relax this requirement [4]. For a fixed CNN $\mathcal{R}_{\boldsymbol{w}}$, the DBP model is a special case of (4): $\hat{\boldsymbol{x}}_{\tilde{\boldsymbol{w}}} \equiv \mathcal{F}_{\tilde{\boldsymbol{w}}}(\boldsymbol{y}; \boldsymbol{A}, \epsilon)$, where $\tilde{\boldsymbol{w}} = (\boldsymbol{w}, \rho)$ are the network parameters, and the network uses measurements together with knowledge of the system and noise power to return an estimate of the image.

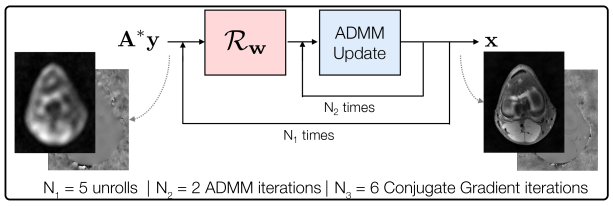

Fig. 1. DBP network architecture, consisting of N_1 unrolled iterations alternating between CNN layers and data consistency layers. The data consistency layer is solved through N_2 unrolled ADMM iterations.

B. Unsupervised Learning

When both input $\{\boldsymbol{y}^{(i)}, \boldsymbol{A}^{(i)}, \epsilon^{(i)}\}_{i=1}^{L}$ and ground-truth $\{\boldsymbol{x}^{(i)}\}_{i=1}^{L}$ training data are available, the network weights can be trained in a traditional end-to-end fashion according to (5). When ground-truth data are not available, we

consider a loss function $\hat{\mathcal{L}}$ imposed in the measurement domain:

$$\min_{\tilde{\boldsymbol{w}}} \frac{1}{L} \sum_{i=1}^{L} \hat{\mathcal{L}} \left(\boldsymbol{A}^{(i)} \hat{\boldsymbol{x}}_{\tilde{\boldsymbol{w}}}^{(i)}, \boldsymbol{y}^{(i)} \right). \tag{12}$$

The measurement loss can be a useful surrogate for the true loss, as the measurements contain (noisy) information about the ground-truth image [9], [14]. Thus, we may hope to learn about the image statistics given a large-enough training set that includes a diversity of measurements.

IV. Methods

We consider the application to under-sampled, multi-channel MRI. The MRI reconstruction task is well-suited to DBP, as the noise statistics are Gaussian and can be measured during a short pre-scan. We first describe the multi-channel MRI forward operator and general sampling strategy. Then we discuss the experimental setup, including the dataset and implementation details.

A. Multi-channel MRI forward operator

In multi-channel MRI, the signal is measured by an array of receive coils distributed around an object, each with a spatially-varying sensitivity profile. In the measurement model, the image is linearly mixed with each coil sensitivity profile, Fourier transformed, and sampled. We can describe the measurement model as $\boldsymbol{A} = \left[(\boldsymbol{PFS}_1)^\top \quad \cdots \quad (\boldsymbol{PFS}_C)^\top \right]^\top \in \mathbb{C}^{M \times N}$, where C is the number of receive coils, $\boldsymbol{S}_c \in \mathbb{C}^{N \times N}$ is a diagonal operator containing the spatial sensitivity profile of the cth coil along the diagonal, $\boldsymbol{F}$ is the Fourier transform operator, and $\boldsymbol{P} \in \{0,1\}^{\frac{M}{C} \times N}$ is a diagonal operator that selects the sampled frequencies.

B. Experimental setup

Data: We used the "Stanford Fully Sampled 3D FSE Knees" dataset from mridata.org, containing 3D Cartesian proton-density knee scans of 20 healthy volunteers. Each 3D volume consisted of 320 slices with matrix size 320×256 and was scanned with an 8-channel receive coil array. Although each slice is fully sampled, in practice the "ground-truth" data itself has noise. To aid in experimental comparison, "noise-free" ground-truth data were created by averaging the data from seven adjacent slices. For each slice, the spatial sensitivity profiles of each coil were estimated using ESPIRiT [15], a self-calibrated parallel imaging method. Ground-truth images were reconstructed by solving (2) using the fully sampled data without regularization. Each slice was then passed through the forward model and retrospectively under-sampled using a different variable-density Poisson-disc sampling pattern [7], [16] with a 16×16 calibration region and acceleration factor $R \approx 12$. Slices from the first 16 volunteers were used for training, discarding the first and last 20 edge slices of each volume (4,384 slices). Similarly, slices from the next two volunteers were used for validation (548 slices), and slices from the last two volunteers were used for testing (548 slices). We added complex-valued Gaussian noise with standard deviation $\sigma = 0.01$ to the noise-free, averaged data.

Implementation: For all experiments we used a Euclidean norm loss function for training. When training with ground-truth (supervised), the loss was applied in the image domain. For unsupervised training, the loss was applied in the measurement (Fourier) domain. We used a U-Net architecture [17] for the CNN auto-encoder, with separate input channels for the real and imaginary components. The U-Net consisted of three encoding layers with ReLU activation functions and 64, 128, and 256 channels, respectively, followed by three similar decoding layers. A final convolutional layer with no activation function mapped the decoder back to two channels. All convolutions used a 3×3 kernel size. For comparison, MoDL [6] was also implemented using the same unrolled parameters and CNN architecture. All networks were implemented in PyTorch.

Evaluation: DBP was separately trained with and without ground-truth data. We also trained MoDL with ground-truth data. In addition, we also evaluated parallel imaging and compressed sensing (PICS) [7] using BART [16], with ℓ_1-Wavelet regularization parameter optimized over the validation set. Normalized root mean-squared error (NRMSE) was used to compare reconstructions.

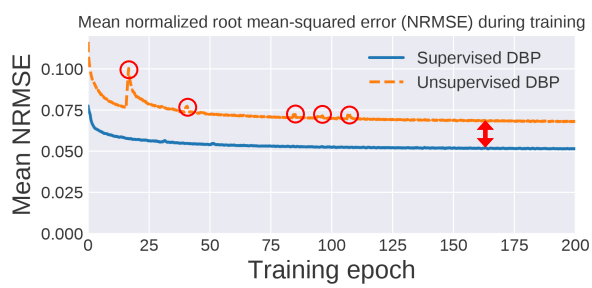

Fig. 2. Mean NRMSE on training set for supervised and unsupervised DBP. Unsupervised DBP has a performance drop (red double-arrow), and noisier updates (red circles).

V. Results

Fig. 2 shows the mean NRMSE on the training set for each epoch. In addition to a performance gap between supervised and unsupervised learning, unsupervised DBP has noisier updates, likely because the loss function in the measurement domain is a noisy surrogate to the NRMSE. Fig. 3 shows the NRMSE across the validation set for different numbers of unrolls during inference. Even though the networks were trained with

5 unrolls, best performance is seen for different number of unrolls (6, 10 and 12 for MoDL, unsupervised DBP, and supervised DBP, respectively). Compared to MoDL, the DBP formulation behaves more stably as the number of unrolls increases, which may be due to the hard data consistency constraint.

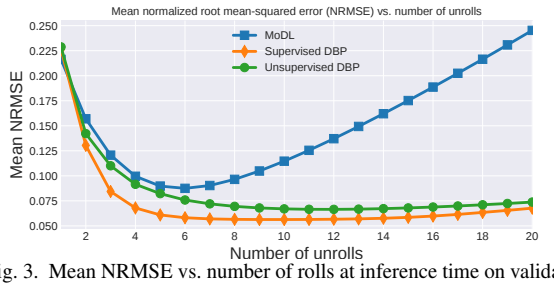

Fig. 3. Mean NRMSE vs. number of rolls at inference time on validation set. All networks were trained with 5 unrolls.

Fig. 4 shows a box plot of the NRMSE on the test set for the different networks and for PICS. Both 5 unrolls and the validation-optimized number of unrolls are shown. At the optimal number of unrolls, unsupervised DBP outperforms PICS.

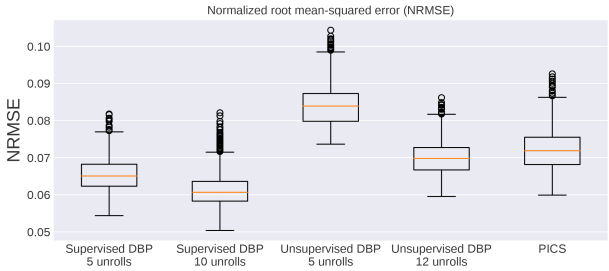

Fig. 4. Box plot of test set NRMSE for supervised and unsupervised DBP at two different unrolls – the first matching unrolls at training, and the second chosen to minimize validation set mean NRMSE. Also shown is PICS NRMSE for optimized regularization on validation set.

Fig. 5 shows some of the intermediate output stages for the supervised and unsupervised DBP networks, indicating that similar structure is learned in both CNNS; however, the supervised DBP appears to better amplify and denoise features in the image. The magnitude reconstructions and error maps of a representative slice from the test set are shown in Fig. 6. Supervised DBP achieves the lowest error, followed by unsupervised DBP, PICS, and MoDL. Small details in edge sharpness are retained with DBP, but reduced with MoDL and PICS.

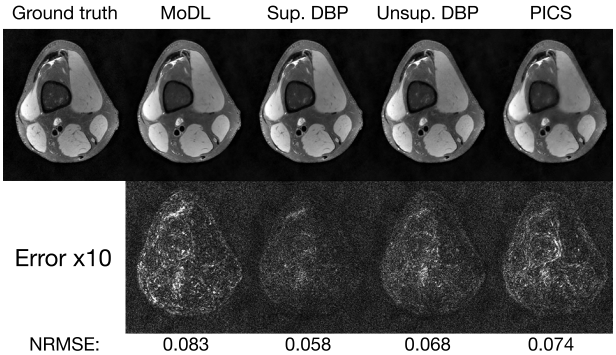

Fig. 6. Comparison of different reconstruction methods to ground-truth image from the test set.

VI. Discussion and Conclusion

There are strong connections to iterative optimization and unrolled deep networks [10], [18], [19]. Jointly optimizing over the images and weights could be viewed a non-linear extension to dictionary learning. Nonetheless, there is a cost in reconstruction error when moving to unsupervised learning, highlighting the importance of a large training data set to offset the missing ground-truth information. The choice of measurement loss function and data SNR may also greatly impact the quality. Fortunately, in many practical settings there is an abundance of under-sampled or corrupted measurement data available for training.

In conclusion, the combination of basis pursuit denoising and deep learning can take advantage of under-sampled data and provide a means to learn model-based deep learning reconstructions without access to ground-truth images.

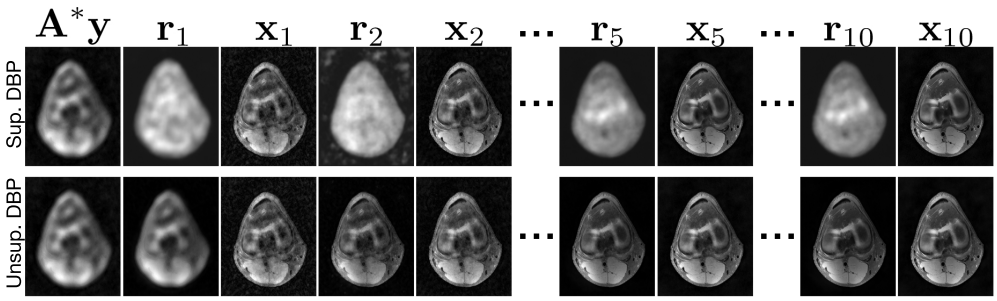

Fig. 5. Output images after different layers in the unrolled network: $\boldsymbol{r}_k$ and $\boldsymbol{x}_k$ are the kth application of the CNN and the data consistency updates, respectively. The supervised CNN amplifies and denoises features to a greater effect compared to the unsupervised CNN.

REFERENCES

[1] J. I. Tamir, S. X. Yu, and M. Lustig, "Unsupervised deep basis pursuit: Learning reconstruction without ground-truth data," in *Proc. Intl. Soc. Mag. Reson. Med 27*, vol. 27, Montreal, May 2019, p. 0660.

[2] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. USA: Omnipress, 2010, pp. 399–406. [Online]. Available: http://dl.acm.org/citation.cfm?id=3104322.3104374

[3] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, "Unrolled optimization with deep priors," *arXiv [cs.CV]*, May 2017. [Online]. Available: http://arxiv.org/abs/1705.08041

[4] 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 Trans Med Imaging*, vol. 37, no. 2, pp. 491–503, 2017.

[5] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data." *Magn Reson Med*, vol. 79, no. 6, pp. 3055–3071, 2017.

[6] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: model based deep learning architecture for inverse problems," *Ieee T Med Imaging*, vol. PP, no. 99, pp. 1–1, 2018.

[7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging." *Magn Reson Med*, vol. 58, no. 6, pp. 1182–95, 2007.

[8] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[9] J. Lehtinen, J. Munkberg, J. Hasselgren, and L. S. preprint arXiv ..., "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv ...*, 2018.

[10] F. Ong, "Low dimensional methods for high dimensional magnetic resonance imaging," 2018.

[11] S. Chen, D. Donoho, and S. M. review, "Atomic decomposition by basis pursuit," *SIAM review*, 2001.

[12] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction - Does AI come at a cost?" *arXiv e-prints*, p. arXiv:1902.05300, Feb 2019.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[14] J. Batson and L. Royer, "Noise2Self: Blind Denoising by Self-Supervision," *arXiv e-prints*, p. arXiv:1901.11365, Jan 2019.

[15] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "Espirit—an eigenvalue approach to autocalibrating parallel mri: Where sense meets grappa," *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.24751

[16] M. Uecker and J. I. Tamir, "bart: version 0.4.04," 2018.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[18] V. Papyan, Y. Romano, and E. M. of Research, "Convolutional neural networks analyzed via convolutional sparse coding," *The Journal of Machine Learning Research*, 2017.

[19] J. Sulam, V. Papyan, Y. Romano, and M. Elad, "Multilayer convolutional sparse modeling: Pursuit and dictionary learning," *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4090–4104, Aug 2018.