MONOTONICALLY CONVERGENT REGULARIZATION BY DENOISING

Yuyang Hu[1], *Jiaming Liu*[1], *Xiaojian Xu*[2], *and Ulugbek S. Kamilov*[1,2]

[1] Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130

[2] Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130

ABSTRACT

Regularization by denoising (RED) is a widely-used framework for solving inverse problems by leveraging image denoisers as image priors. Recent work has reported the state-of-the-art performance of RED in a number of imaging applications using pre-trained deep neural nets as denoisers. Despite the recent progress, the stable convergence of RED algorithms remains an open problem. The existing RED theory only guarantees stability for convex data-fidelity terms and nonexpansive denoisers. This work addresses this issue by developing a new *monotone RED (MRED)* algorithm, whose convergence does not require nonexpansiveness of the deep denoising prior. Simulations on image deblurring and compressive sensing recovery from random matrices show the stability of MRED even when the traditional RED diverges.

Index Terms— Inverse problems, plug-and-play priors, regularization by denoising, model-based deep learning.

1. INTRODUCTION

Many imaging problems, such as deblurring and super-resolution, can be formulated as *inverse problems* involving the recovery of an image $\boldsymbol{x} \in \mathbb{R}^n$ from its measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}, \qquad (1)$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the measurement operator and $\boldsymbol{e} \in \mathbb{R}^n$ is the noise. Since many inverse problems are ill-posed, it is common to formulate the solution as a *regularized inversion*, expressed as an optimization problem

$$\widehat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) \quad \text{with} \quad f(\boldsymbol{x}) = g(\boldsymbol{x}) + h(\boldsymbol{x}), \qquad (2)$$

where g a *data-fidelity term* that quantifies consistency with the observed measurements $\boldsymbol{y}$ and h is the *regularizer* that enforces prior knowledge on $\boldsymbol{x}$. For example, when the noise vector $\boldsymbol{e}$ in (1) is the *additive white Gaussian noise (AWGN)*, the popular data-fidelity term is the least-squares function $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$. On the other hand, many widely-used image regularizers are based on sparsity promoting functions

---

of the form $h(\boldsymbol{x}) = \tau \|\boldsymbol{W}\boldsymbol{x}\|_1$, where $\tau > 0$ is the regularization parameter and $\boldsymbol{W}$ is a suitable transform [1–3].

There has been considerable recent interest in leveraging image denoisers as priors for the recovery of $\boldsymbol{x}$ [4, 5]. One of the most popular frameworks in this context is *regularization by denoising (RED)* [6], which uses a denoiser $\mathsf{D} : \mathbb{R}^n \to \mathbb{R}^n$ as a regularizer within an iterative algorithm. For example, the *steepest descent* variant of RED (RED-SD) [6] is implemented by iterating the following steps until convergence

$$\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma \mathsf{G}(\boldsymbol{x}^{k-1}) \quad \text{with} \qquad (3a)$$

$$\mathsf{G}(\boldsymbol{x}^{k-1}) := \nabla g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{D}(\boldsymbol{x}^{k-1})), \quad (3b)$$

where $\gamma > 0$ is the step size, and $\tau > 0$ is the regularization parameter. For a locally homogeneous D that is nonexpansive and has a symmetric Jacobian, it can be shown [6,7] that RED algorithms solve (2) with a regularizer

$$h(\boldsymbol{x}) = \frac{\tau}{2}\boldsymbol{x}^\mathsf{T}(\boldsymbol{x} - \mathsf{D}(\boldsymbol{x})). \qquad (4)$$

The state-of-the-art performance of RED on many computational imaging problems has motivated a great deal of research on RED and its applications [8–13].

Despite the recent activity around the topic, the stable convergence of RED algorithms remains an open problem from both theoretical and practical perspectives. The existing theoretical guarantees for the convergence of RED algorithms require the denoiser D to be a *nonexpansive* operator [6, 7, 9, 11, 13]. On the other hand, it has been empirically observed that RED algorithms can easily diverge for expansive denoisers (see, for example, Table II in [9]). With the explosion of interest in using deep neural networks within iterative algorithms, there is consequently a pressing need for more flexible strategies for desining RED algorithms that are stable for nonconvex functions g and expansive operators D.

This paper addresses the need for stable RED algorithms by proposing a new *monotone RED (MRED)* algorithm that can offer stable convergence for nonconvex data-fidelity terms and expansive deep image denoisers. MRED defines and uses an explicit loss based on the fixed-point interpretation of RED that does not rely on the existence of the explicit RED regularizer h. MRED extends the heuristic *backtracking line search (BLS)* strategy proposed in the prior work [12]

Algorithm 1 RED with BLS

1: **input:** $\boldsymbol{x}^0 \in \mathbb{R}^n; \gamma, \varepsilon > 0; \beta \in (0,1)$
2: **for** $k = 1, 2, \ldots, t$ **do**
3:     $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\mathsf{G}(\boldsymbol{x}^{k-1})$
4:     **while** $\|\mathsf{G}(\boldsymbol{x}^k)\|_2 > \|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2$ **do**
5:         $\gamma \leftarrow \beta\gamma$
6:         $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\mathsf{G}(\boldsymbol{x}^{k-1})$
7:         **if** $\gamma < \varepsilon$ **then**
8:             **return:** $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^{k-1}$
9: **output:** $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^t$

Algorithm 2 Monotone RED (MRED)

1: **input:** $\boldsymbol{x}^0 \in \mathbb{R}^n; \gamma, \alpha_0, \varepsilon > 0; \beta \in (0,1); \theta \in (0, 1/2)$
2: **for** $k = 1, 2, \ldots, t$ **do**
3:     $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\mathsf{G}(\boldsymbol{x}^{k-1})$
4:     $\alpha \leftarrow \alpha_0$
5:     **while** $\varphi(\boldsymbol{x}^k) > \varphi(\boldsymbol{v}^{k-1}) - \alpha\theta\|\nabla\varphi(\boldsymbol{x}^{k-1})\|_2^2$ **do**
6:         $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \alpha\nabla\varphi(\boldsymbol{x}^{k-1})$
7:         $\alpha \leftarrow \beta\alpha$
8:         **if** $\alpha < \varepsilon$ **then**
9:             **return:** $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^{k-1}$
10: **output:** $\boldsymbol{x}^* \leftarrow \boldsymbol{x}^t$

by adopting a more principled BLS strategy. MRED is the first RED method which is guaranteed to converge for any denoiser—irrespective of its expansiveness. Our simulations on image deblurring and compressive sensing from random matrices illustrate the stability of MRED compared to the traditional RED-SD with and without BLS.

2. BACKGROUND

There has been a broad interest in methods that integrate deep neural nets into iterative algorithms for solving imaging problems. One of the earliest methods is *plug-and-play priors (PnP)* [4], which proposed to extend the traditional proximal optimization [14] by replacing the proximal operator with a more general denoiser. It has been shown that the combination of proximal algorithms with advanced denoisers, such as BM3D [15] or DnCNN [16], leads to the state-of-the-art performance for various imaging problems. Remarkably, the heuristic of using denoisers within iterative algorithms exhibited great empirical success [13, 17–20] and inspired a significant follow up work on the so-called *model-based deep learning* method that include RED, *denoising-based approximate message passing (D-AMP)*, *deep unfolding (DU)*, and *deep equilibrium models (DEQ)* [21–26].

One of the key motivations of RED is its formulation as an optimization problem (2) with an explicit regularizer h in (4). However, the existence of h only holds under a set of assumptions on D, such as its *nonexpansiveness* [6, 7, 11]. As a reminder, an operator D is nonexpansive, if it satisfies

$$\|\mathsf{D}(\boldsymbol{x}) - \mathsf{D}(\boldsymbol{z})\|_2 \leq \|\boldsymbol{x} - \boldsymbol{z}\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n. \quad (5)$$

Remarkably, nonexpansiveness is also used in the convergence analysis of RED algorithms when interpreting them as fixed-point iterations seeking zeros of the operator G [9, 13]. While it is possible to train nonexpansive deep neural nets via spectral normalization [9, 20, 27], nonexpansiveness can hurt the performance and reduce the flexibility to use existing expansive denoisers. It is worth noting that nonexpansiveness, as well as more restrictive conditions *firm nonexpansiveness* and *contractiveness*, are used to ensure convergence of other related frameworks, such as PnP and DEQ [9, 26]. Thus,

Fig. 1. The six test images used in our numerical evaluation. From left to right, *baby, bird, butterfly, leaves, starfish, head*.

while our focus is on RED, our approach can be applied to develop more stable variants of other related frameworks.

3. PROPOSED METHOD

3.1. RED with Backtracking Line Search

RED algorithms can be interpreted as fixed-point iterations seeking a vector $\boldsymbol{x}^*$ in the zero set of the operator G in (3b)

$$\boldsymbol{x}^* \in \mathrm{zer}(\mathsf{G}) := \{\boldsymbol{x} \in \mathbb{R}^n : \mathsf{G}(\boldsymbol{x}) = \boldsymbol{0}\}. \quad (6)$$

This implies that the goal of a RED algorithm is to reduce the quantity $\|\mathsf{G}(\boldsymbol{x}^k)\|_2$ at every iteration $k \geq 1$. This observation is behind the *backtracking line search (BLS)* strategy for RED summarized in Algorithm 1. The main idea behind RED (BLS) is to shrink the step size $\gamma > 0$ by a factor $0 < \beta < 1$ whenever the distance to $\mathrm{zer}(\mathsf{G})$ is not improving. The algorithm stops when the step-size becomes smaller than a pre-defined quantity $\varepsilon > 0$ or when the maximum number of iterations $t \geq 1$ is reached. This strategy was originally proposed in [12] as an heuristic to stabilize the convergence of *regularization by artifact removal (RARE)* variant of RED that replaces image denoisers by artifact-removal operators. However, the convergence of the algorithm was never formally compared to the traditional RED with a fixed step-size.

When the data-fidelity term g is a convex Lipschitz continuous function and the denoiser is nonexpansive, one can show that there exists a small enough step size so that the iterates of RED (BLS) satisfy $\|\mathsf{G}(\boldsymbol{x}^k)\|_2 \leq \|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2$ (see the analysis in [9]). However, for nonconvex data-fidelity terms and/or expansive denoisers there is no guarantee for the existence of such a step size, which means that RED (BLS) might terminate prematurely.
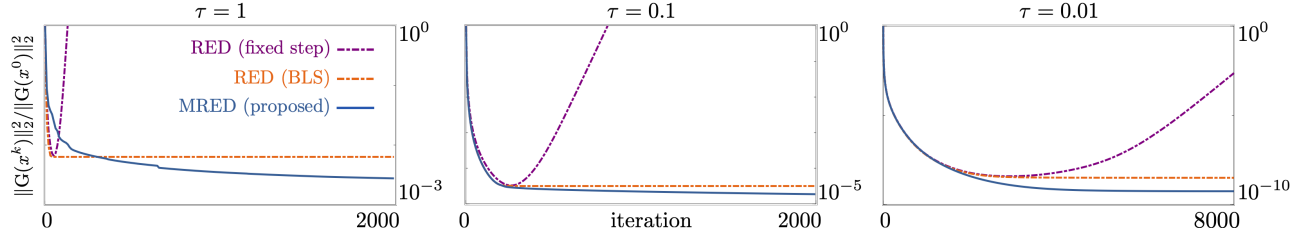
Fig. 2. Illustration of the MRED convergence for image deblurring for different regularization parameters $\tau \in \{1, 0.1, 0.01\}$. The figures plot the average normalized distance to $\mathrm{zer}(\mathsf{G})$ over the six test images against the iteration number. We compare three algorithms, the traditional RED with a fixed step-size parameter, RED with BLS, and the proposed MRED. Note how MRED achieves the smallest distance to $\mathrm{zer}(\mathsf{G})$ for all three plots.
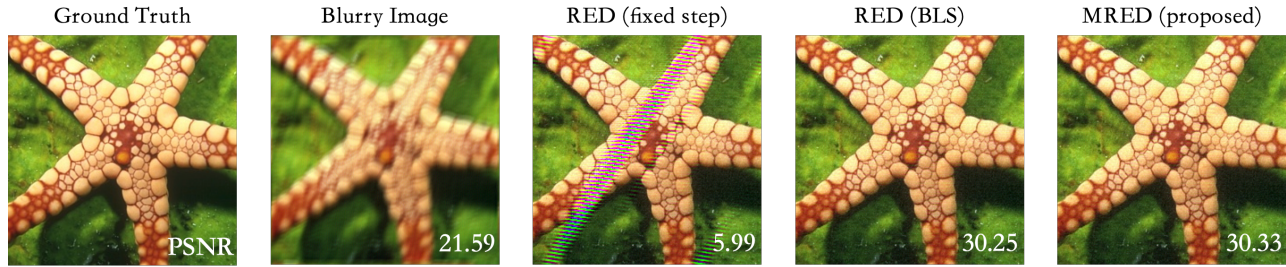

Fig. 3. Visual results for image deblurring with a 17×17 blur kernel from [28] and $30\,\mathrm{dB}$ AWGN noise on the *Starfish* image. We ran three algorithms the traditional RED with the fixed step-size, RED (BLS), and MRED for the regularization parameter $\tau = 0.01$. Note the poor visual quality of the RED image with the fixed step-size due to divergence of the algorithm.

3.2. Monotone RED

The proposed *monotone RED (MRED)* method, summarized in Algorithm 2, addresses this issue by reformulating the search for a vector in $\mathrm{zer}(\mathsf{G})$ as an optimization problem. Consider the loss function

$$\varphi(\boldsymbol{x}) := \frac{1}{2}\|\mathsf{G}(\boldsymbol{x})\|_2^2, \tag{7}$$

which measures the squared distance of a vector to $\mathrm{zer}(\mathsf{G})$. It is clear that whenever $\mathrm{zer}(\mathsf{G}) \neq \varnothing$, the minimizers of φ are vectors in $\mathrm{zer}(\mathsf{G})$. Additionally, whenever φ is a smooth function, we can use the traditional optimization theory to perform BLS (see Secion 9.2 in [29]). To that end, let φ be a smooth function that has a Lipschitz continuous gradient with constant $L > 0$ and consider the gradient update

$$\boldsymbol{x}^+ \leftarrow \boldsymbol{x} - \alpha \nabla \varphi(\boldsymbol{x}), \tag{8}$$

where we will assume that the step satisfies $0 < \alpha \leq 1/L$. Note that one can use the automatic differentiation capability of modern deep learning frameworks to evaluate $\nabla \varphi$. Then, from the Lipschitz continuity of the gradient

$$\begin{aligned}
\varphi(\boldsymbol{x}^+) &\leq \varphi(\boldsymbol{x}) + \nabla\varphi(\boldsymbol{x})^\mathsf{T}(\boldsymbol{x}^+ - \boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{x}^+ - \boldsymbol{x}\|_2^2 \\
&\leq \varphi(\boldsymbol{x}) - \frac{\alpha}{2}\|\nabla\varphi(\boldsymbol{x})\|_2^2 \\
&\leq \varphi(\boldsymbol{x}) - \alpha\theta\|\nabla\varphi(\boldsymbol{x})\|_2^2,
\end{aligned} \tag{9}$$

with $0 < \theta < 1/2$, where in the second row we used (8) and the fact that $0 < \alpha \leq 1/L$. Therefore, there always exists a small enough step-size $\alpha > 0$ such that (9) is satisfied. MRED uses the violation of the condition (9) as an indicator to reduce its step-size parameter. It is worth mentioning that for $\theta = 0$, the condition in Line 5 of MRED is equivalent to the condition of the while loop in Line 4 of RED (BLS). However, unlike RED (BLS), MRED guarantees the existence of a small enough step-size even for nonconvex functions g and expansive denoisers D, so long the function φ is smooth.

MRED is a hybrid algorithm in the sense that it uses the traditional RED update (Line 3 of MRED) whenever it can reduce the distance to $\mathrm{zer}(\mathsf{G})$ and switches to the gradient update (Line 6 of MRED) if this is not possible. The benefit of the proposed hybrid approach is that it maintains the fast fixed-point convergence of the traditional RED algorithm whenever possible. Additionally, the gradient $\nabla\varphi$ needs to be evaluated only when the traditional RED update cannot reduce the loss, thus limiting the computational overhead relative to the traditional RED algorithm. In short, MRED maintains all the benefits of the traditional RED algorithm, while also providing additional flexibility and stability. In Section 4, we validate the effectiveness of MRED by comparing it both with the traditional RED with a fixed step-size and RED (BLS) on two different types of inverse problems.
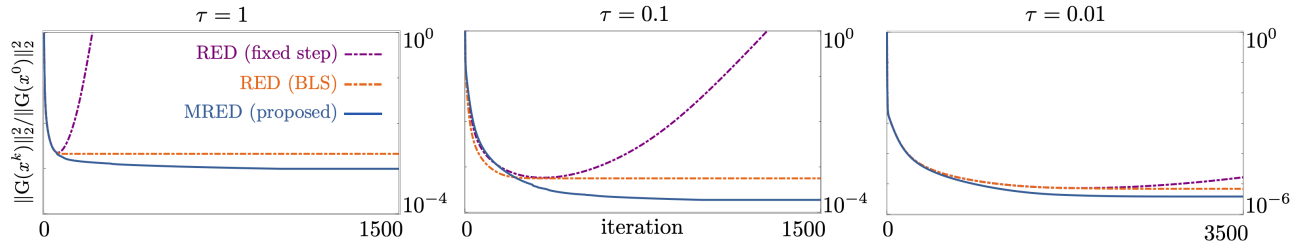
428

Fig. 4. Illustration of the MRED convergence for compressive sensing from random measurements for different regularization parameters $\tau \in \{1, 0.1, 0.01\}$. The figures plot the average normalized distance to $\mathrm{zer}(\mathsf{G})$ over the six test images against the iteration number. We compare three algorithms, the traditional RED with a fixed step-size parameter, RED with BLS, and the proposed MRED. Note how MRED achieves the best convergence behaviour for all three plots.

4. NUMERICAL RESULTS

In this section, we numerically illustrate the ability of MRED to stabilize the convergence for expansive deep denoising priors. We consider two scenarios: (a) image deblurring and (b) compressive sensing using random projections. Our deep neural net prior is the simplified Unet architecture [30], obtained by removing all the batch-normalization layers. We trained the denoiser as a residual network $\mathsf{R} := \mathsf{I} - \mathsf{D}$ that predicts the noise residual from a noisy input, without using any spectral normalization for controlling the nonexpansiveness of the network. Hence, our network is expected to be expansive. The training data was generated by adding AWGN to images from the BSD500 [31] and DIV2K [32] datasets, by cropping the images into small patches of 96×96 pixels with stride 10. Note that the gradient in (8) is given by the expression $\nabla \varphi(\boldsymbol{x}) = \boldsymbol{A}^{\mathsf{T}} \boldsymbol{A} \mathsf{G}(\boldsymbol{x}) + \tau \nabla \mathsf{R}(\boldsymbol{x})^{\mathsf{T}} \mathsf{G}(\boldsymbol{x})$, where the last term was computed by using the autodiff functionality of PyTorch. The normalized distance $\|\mathsf{G}(\boldsymbol{x}^k)\|_2^2 / \|\mathsf{G}(\boldsymbol{x}^0)\|_2^2$ was used to quantify the fixed-point convergence of three RED algortihms: (a) the traditional RED with a fixed step size; (b) RED with BLS in Algorithm 1; and (c) MRED in Algorithm 2. Following the theoretical analysis in [9], we fix the step-size of the traditional RED $\gamma = 0.8$, which is small enough to theoretically ensure convergence under nonexpansive denoisers. Note that we fixed The quantity $\|\mathsf{G}(\boldsymbol{x}^k)\|_2^2 / \|\mathsf{G}(\boldsymbol{x}^0)\|_2^2$ is expected to approach zero as the algorithms converge to $\mathrm{zer}(\mathsf{G})$. All the quantitative results were obtained by averaging over six test images shown in Fig. 1.

4.1. Image Deblurring

The measurement model in image deblurring with uniform blur can be expressed as in (1), where the matrix $\boldsymbol{A}$ denotes a two-dimensional convolution between the clean image $\boldsymbol{x}$ and the blur kernel. We consider the blur kernel of size 17×17 and the AWGN vector $\boldsymbol{e}$ corresponding to the input SNR of $30\,\mathrm{dB}$. Fig. 2 illustrates the convergence behaviour of all three RED algorithms. First, note the divergence of the traditional RED with the fixed step-size, which is not surprising since

our CNN prior has been trained without any Lipschitz constraints. Note how RED (BLS) stabilizes the convergence, but stops making any progress prematurely due to the fact that it cannot find a small enough step-size to decrease the distance to $\mathrm{zer}(\mathsf{G})$. Finally, as expected, MRED achieves the best convergence behaviour for all three values of τ. Fig. 3 shows some visual examples of the performance of all three algorithms, where RED yields a suboptimal image due to its divergence and MRED achieves the highest quality image.

4.2. Compressive Sensing

We consider the traditional compressive sensing measurement setup $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is orthogonalized version of a matrix with elements that are i.i.d. zero-mean Gaussian random variables of variance of $1/m$. We set the compression ratio m/n to be 0.1 and consider a similar setting to image deblurring by running all three algorithms for three values of τ: 1, 0.1, and 0.01. As before, our baseline methods are the traditional RED with a fixed step-size and RED (BLS). Fig. 4 illustrates the divergence of the traditional RED and convergence of MRED for all the regularization parameters τ.

5. CONCLUSION

There is a growing interest in iterative algorithms, such as PnP, RED, or DEQ, for solving inverse problems by using deep neural nets as priors. While such algorithms have often lead to significant improvements in the imaging quality, they have also been shown to suffer from convergence issues when the deep neural net is expansive. In this paper, we proposed a new MRED algorithm that uses a novel line search strategy to ensure monotonic convergence of the steepest descent variant of RED (RED-SD) even for nonconvex data-fidelity terms and expansive denoisers. Our numerical results highlight the stability of MRED relative to the traditional RED-SD algorithm, which diverges for our expansive denoiser. While our focus was on RED, the future work will look into the potential of our strategy to be applicable to other related frameworks such as PnP and DEQ.

6. REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, November 1992.

[2] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, December 2007.

[3] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[4] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Process. and INf. Process. (GlobalSIP)*, 2013, pp. 945–948.

[5] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1043–1055, Feb. 2015.

[6] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[7] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, Mar. 2019.

[8] C. A. Metzler, P. Schniter, A. Veeraraghavan, and R. G. Baraniuk, "prDeep: Robust phase retrieval with a flexible deep network," in *Proc. 35th Int. Conf. Machine Learning (ICML)*, Stockholm, Sweden, June 2018, pp. 3501–3510.

[9] Y. Sun, J. Liu, and U. S. Kamilov, "Block coordinate regularization by denoising," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 908–921, 2020.

[10] G. Mataev, M. Elad, and P. Milanfar, "DeepRED: Deep image prior powered by RED," in *Proc. IEEE Int. Conf. Comp. Vis. Workshops (ICCVW)*, Seoul, South Korea, Oct 27-Nov 2, 2019, pp. 1–10.

[11] R. Cohen, M. Elad, and P. Milanfar, "Regularization by denoising via fixed-point projection (RED-PRO)," *SIAM J. Imaging Sci.*, vol. 14, no. 3, pp. 1374–1406, 2021.

[12] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "RARE: Image reconstruction using deep priors learned without ground truth," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1088–1099, 2020.

[13] J. Liu, S. Asif, B. Wohlberg, and U. S. Kamilov, "Recovery analysis for plug-and-play priors using the restricted eigenvalue condition," in *Proc. Advances in Neural Information Processing Systems 35*, December 6-14, 2021.

[14] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2014.

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," in *Proc. SPIE Elctron. Imag.*, San Jose, CA, January 2008, vol. 6812, p. 681207.

[16] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, July 2017.

[17] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comp. Imag.*, vol. 3, no. 1, pp. 84–98, March 2017.

[18] Y. Sun, B. Wohlberg, and U. S. Kamilov, "An online plug-and-play algorithm for regularized image reconstruction," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 395–408, Sept. 2019.

[19] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "A convergent image fusion algorithm using scene-adapted Gaussian-mixture-based denoising," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 451–463, Jan. 2019.

[20] E. K. Ryu, J. Liu, S. Wnag, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, Long Beach, CA, USA, June 2019, pp. 5546–5557.

[21] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Advances in Neural Information Processing Systems 29*, 2016, pp. 10–18.

[22] C. A. Metzler, A. Mousavi, and R. G. Baraniuk, "Learned D-AMP: Principled neural network based compressive image recovery," in *Proc. Advances in Neural Information Processing Systems 31*, Long Beach, CA, USA, Dec. 2017.

[23] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimizationinspired deep network for image compressive sensing," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 1828–1837.

[24] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.

[25] C Bertocchi, E. Chouzenoux, M.-C. Corbineau, J.-C. Pesquet, and M. Prato, "Deep unfolding of a proximal interior point method for image restoration," *Inverse Probl.*, vol. 36, no. 3, pp. 034005, 2020.

[26] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1123–1133, Oct. 2021.

[27] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *International Conference on Learning Representations (ICLR)*, Vancouver, Canada, Apr. 2018.

[28] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman, "Understanding and evaluating blind deconvolution algorithms," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1964–1971.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.

[30] 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, pp. 234–241.

[31] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, Vancouver, Canada, July 7-14, 2001, pp. 416–423.

[32] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE CVPR WKSHP*, 2017, pp. 126–135.