ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10446385

VARIATIONAL ANALYSIS OF ADVERSARIAL REGULARIZATION FOR SOLVING INVERSE PROBLEMS

Abhishek Shreekant Bhandiwad[1], *Abijith Jagannath Kamath*[2], *Siddarth Asokan*[3] and *Chandra Sekhar Seelamantula*[2]

[1]Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru 560 012
[2]Department of Electrical Engineering, Indian Institute of Science, Bengaluru 560 012
[3] Microsoft Research Lab India, Bengaluru 560 001
Email:{abhishekbs, abijithj, css}@iisc.ac.in, sasokan@microsoft.com

ABSTRACT

Inverse problems form the backbone of modern signal/image processing and computational imaging, where signal reconstruction from corrupted measurements follows an optimization problem. The objective function is the sum of a data-fidelity term and a regularization functional that enforces desired properties in the reconstruction. The adversarial regularization (AR) framework is an unsupervised, data-driven approach for solving inverse problems, where the regularization function is learnt adversarially as a critique between the ground-truth distribution and the distribution of unregularized reconstructions. Thereafter, the solution to the regularized inverse problem follows an iterative technique. In this paper, we analyze the AR framework from a variational perspective, and, using Euler-Lagrange conditions, obtain the optimal regularization function in closed-form. The overall objective function is smooth and readily amenable to gradient descent minimization. We introduce momentum into the iterates as a natural extension to accelerate convergence. Since the optimal solutions are obtained in closed-form, our approach to solving inverse problems does not require prior training whilst being data-driven. We demonstrate the proposed technique on image deconvolution and show that the reconstruction performance of the proposed techniques measured in terms of peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM) are identical to the learnt counterparts.

Index Terms— Inverse problems, image restoration, adversarial regularization, calculus of variations, Euler-Lagrange analysis.

1. INTRODUCTION

Several signal/image processing and computational imaging tasks involve solving inverse problems, where the goal is to reconstruct/recover/restore the ground-truth signal or (vectorized) image $\mathbf{x} \in \mathbb{R}^n$ from noisy measurements $\mathbf{y} \in \mathbb{R}^m$ of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{H}$ denotes the *forward operator* and $\mathbf{w} \in \mathbb{R}^m$ denotes additive noise. Several applications such as denoising, deconvolution, sparse signal restoration, compressed image reconstruction and low-rank plus sparse decomposition are *linear inverse problems* where the forward operator $\mathbf{H} \in \mathbb{R}^{m\times n}$ is a linear transformation [1, 2]. Typically, with $m \leq n$, linear inverse problem may be ill-posed requiring additional constraints on the solution space, sparsity being the most popular. Contemporary techniques [3,4] for solving inverse problems involve solving an optimization problem of the type

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\left\|\mathbf{H}\mathbf{x}-\mathbf{y}\right\|_2^2 + \lambda f(\mathbf{x}). \tag{P}$$

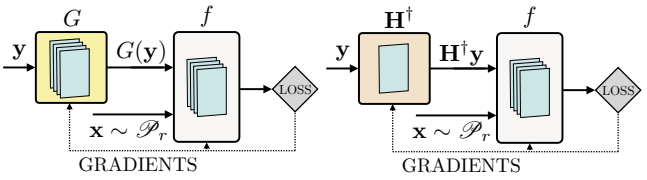
The objective function is the sum of the convex data-fidelity loss $\frac{1}{2}\left\|\mathbf{H}\mathbf{x}-\mathbf{y}\right\|_2^2$, and a regularization functional $f : \mathbb{R}^n \to \mathbb{R}$ that penalizes solutions that are not in the desired constraint space. $\lambda > 0$ is called the *regularization parameter* that trades-off between the data-fidelity and regularization terms. From a statistical perspective, the regularization functional approximates the log-density of the prior on the desired solutions. Examples of model-based design of the regularization functionals for image restoration tasks include (a) Tikhonov regularization $f(\mathbf{x}) = \left\|\mathbf{L}\mathbf{x}\right\|_2^2$, where $\mathbf{L}$ is a regularization operator [5], (b) wavelet thresholding $f(\mathbf{x}) = \left\|\Psi\mathbf{x}\right\|_1$, where Ψ denotes the dyadic wavelet transform [6], and (c) total-variation $f(\mathbf{x}) = \left\|\mathbf{D}\mathbf{x}\right\|_1$, where $\mathbf{D}$ denotes the gradient operator [7]. Typically, model-based regularization functionals rely on sparsity in the ground-truth signal and are application-specific.

Q1: Can one learn optimal data-driven regularization functionals using training examples for a specific task?

With the introduction of deep learning, several learning-based approaches for solving inverse problems are available in the literature (see [2, 8–12] and references therein). Several frameworks have been developed that combine deep learning with model-based techniques. These techniques utilize training data to improve reconstruction performance whilst retaining convergence or performance bounds [13–17]. However, learning-based methods often require large amounts of training data, and involve the computationally expensive task of training deep neural networks. The training problems are typically supervised in the sense that they depend on the availability of large training datasets with explicit input-output pairs.

The adversarial-regularization (AR) framework [18] is a data-driven approach for solving inverse problems that do not require explicit input-output pairs. The central idea of the AR framework is to model the regularization functional as a deep neural network optimized using training data. Protracting *Q1*, we also attempt to address the following question in this paper:

Q2: Can one learn optimal data-driven regularization functionals using unsupervised training examples of a specific task, without training?



(a) GAN (b) Adversarial regularisation

Fig. 1. Schematic of training problems in (a) GAN and (b) AR: the generator network in a GAN is replaced by a linear transformation that is fixed to the pseudoinverse $\mathbf{H}^\dagger$ of the forward operator in AR. The variational problem for the discriminator network and the regularization functional are identical.

1.1. Adversarial Regularization

We provide a brief introduction to the adversarial-regularization (AR) framework [18, 19]. In AR, the regularization function f is optimized using training data in an adversarial and unsupervised fashion. Let $\mathbf{x}_r \sim \mathscr{P}_r$ denote the training samples of the ground-truth signals from a distribution $\mathscr{P}_r$, and $\mathbf{y} \sim \mathscr{P}_o$ denote the training samples of the measurements from a distribution $\mathscr{P}_o$, that follow Eq. (1) with a fixed noise level. Naïve or unregularized reconstructions using the measurements alone are denoted as $\mathbf{x}_u = \mathbf{H}^\dagger \mathbf{y} \sim \mathscr{P}_u$, where $\mathbf{H}^\dagger$ denotes the pseudoinverse of $\mathbf{H}$. The aim is to construct regularization functionals $f : \mathbb{R}^n \to \mathbb{R}$ such that the value is high for $\mathbf{x}_u$ and low for $\mathbf{x}_r$ by solving the continuous-domain minimization problem with loss function

$$\mathop{\mathbb{E}}_{X\sim\mathscr{P}_r}[f(X)] - \mathop{\mathbb{E}}_{X\sim\mathscr{P}_u}[f(X)] + \lambda_{\mathrm{GP}} \mathop{\mathbb{E}}_{X\sim\mathscr{P}_c}\left[(\|\nabla f(X)\|_2 - 1)_+^2\right], \quad (2)$$

where $\lambda_{\mathrm{GP}} > 0$ is a hyperparameter, $(x)_+ = \max\{0, x\}$ denotes the rectified linear unit, and $\mathscr{P}_c = \rho\mathscr{P}_r + (1-\rho)\mathscr{P}_u$, for some $\rho \in (0, 1)$, denotes a convex combination of $\mathscr{P}_r$ and $\mathscr{P}_u$. The minimizer in Eq. (2) is an approximation to the maximizer of the Wasserstein distance between $\mathscr{P}_r$ and $\mathscr{P}_u$ [20]

$$\mathrm{Wass}(\mathscr{P}_r, \mathscr{P}_u) = \sup_{f\in\text{Lip-1}} \mathop{\mathbb{E}}_{X\sim\mathscr{P}_u}[f(X)] - \mathop{\mathbb{E}}_{X\sim\mathscr{P}_r}[f(X)],$$

where Lip-1 denotes the family of Lipschitz-1 functionals. The term $\mathop{\mathbb{E}}_{X\sim\mathscr{P}_c}\left[(\|\nabla f(X)\|_2 - 1)_+^2\right]$ in Eq. (2) enforces the norm of the gradient of f to be close to unity, promoting Lipschitz-1 functionals. The sample estimates for the expectation are taken as arbitrary convex combinations of $\mathscr{P}_r$ and $\mathscr{P}_u$. Note that the learning problem does not require explicit signal-measurement pairs, and in that sense, the learning problem is unsupervised. Lunz *et al.* [18] parametrize f as a deep neural network that is learnt using the loss function given in Eq. (2). Once the optimal regularization functional, denoted as f_{AR} is learnt, the optimization problem (**P**) is solved using subgradient descent [21]. Lunz *et al.* [18] show empirically that the learnt regularizer leads to solutions better than benchmark unsupervised techniques. Subsequently, Mukherjee *et al.* [19] solve Eq. (2) with a convexity constraint on the solution that provides theoretical convergence guarantees for the subgradient-descent algorithm.

Relationship between AR and GANs: The AR framework is closely linked to the field of generative modelling using generative adversarial networks (GANs) [22, 23]. The discriminator, parametrized as a neural network in a GAN, is akin to the regularization functional in AR; and the generator, parametrized as a neural network in a GAN, is akin to the pseudoinverse $\mathbf{H}^\dagger$ of the forward model. The training problems in GANs and AR are identical. Fig. 1 shows the learning schematics for a GAN and AR — the regularization functional in AR is equivalent to the discriminator network in a GAN. The optimization in Eq. (2) is considered in the flavour of GANs called Wasserstein GANs (WGANs) [24, 25]. Asokan and Seelamantula [26, 27] improve the performance of WGANs by analyzing the discriminator optimization in the Euler-Lagrange framework [28] — this forms the basis for our paper. The key contribution of this paper within the AR framework is that our iterative reconstruction techniques do not require prior training whilst being data-driven.

1.2. Contributions of This Paper

The main claim of this paper is that adversarial regularization can be achieved without having to train a neural network. We analyze solutions to a variant of Eq. (2) and show, using Euler-Lagrange analysis, that the solution can be represented in closed-form using the family of polyharmonic spline functions (Section 2). The resulting optimal regularization function is continuously differentiable, and its gradient can be computed efficiently using sample estimates. With the optimal regularization function in (**P**), we solve the optimization using gradient descent (AR-ELA). As a natural extension, we introduce Nesterov's momentum in the update steps to accelerate convergence (AR-ELA+) in Section 3. In Section 4, we demonstrate the performance of the proposed techniques for image deconvolution and show that the performance measured in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are comparable to benchmark learning-based techniques. From the perspective of interpretability, our analysis exposes the classes of regularization functions that can be obtained by training neural networks using the AR framework.

2. EULER-LAGRANGE ANALYSIS OF ADVERSARIAL REGULARIZATION

Consider the following variational problem:

$$f^*_{p,n} = \arg\min_{f\in\mathrm{BL}_{p,2}} \mathop{\mathbb{E}}_{X\sim\mathscr{P}_r}[f(X)] - \mathop{\mathbb{E}}_{X\sim\mathscr{P}_u}[f(X)] + \lambda_{\mathrm{GP}} \int_\chi \|\nabla^p f(\mathbf{x})\|_2^2 \,\mathrm{d}\mathbf{x}, \quad (3)$$

where λ_{GP} is a regularization parameter, $p \in \mathbb{N}$, χ denotes the convex hull of the supports of $\mathscr{P}_r$ and $\mathscr{P}_u$, and $\mathrm{BL}_{p,2}$ denotes the Beppo-Levi space of order $p, 2$ [29]. The variational problem in Eq. (3) is a generalization of Eq. (2) and are equivalent when $p = 1$. The objective does not explicitly enforce the minimizer to be Lipschitz-1, however, enforces higher-order smoothness on the minimizer, constraining the solution to lie in the Beppo-Levi space $\mathrm{BL}_{p,2}$. The generalization of the higher-order penalty is popular in image processing tasks, with the higher-order gradient penalty applied directly on the image [30]. The variational problem in Eq. (3) is a familiar problem in the context of GANs [26, 27, 31], which improves the performance of image generation as compared to other flavours of GANs.

Euler-Lagrange analysis: We analyze the solution space of the variational problem in Eq. (3) using the Euler-Lagrange framework [28]. The objective function of the variational problem in Eq. (3) can be written in integral form as:

$$\int_\chi \left(f(\mathbf{x})\,(\mathscr{P}_r(\mathbf{x}) - \mathscr{P}_u(\mathbf{x})) + \lambda_{\mathrm{GP}} \|\nabla^p f(\mathbf{x})\|_2^2 \right) \mathrm{d}\mathbf{x},$$

which is a function of f and its derivatives. Invoking the Euler-Lagrange condition for optimality, it can be shown that the optimal regularization functional $f^*_{p,n}$ satisfies:

$$\Delta^p f^*_{p,n}(\mathbf{x}) = \frac{(-1)^{p+1}}{2\lambda_{\mathrm{GP}}} \left(\mathscr{P}_r(\mathbf{x}) - \mathscr{P}_u(\mathbf{x})\right), \tag{4}$$

where $\Delta = \nabla \cdot \nabla = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ denotes the Laplacian operator, and $\Delta^p \stackrel{\text{def.}}{=} \Delta(\Delta^{p-1})$ denotes the iterated Laplacian operator. The optimality condition in Eq. (4) is a partial differential equation where the particular solution can be computed using the Green's function of Δ^p, i.e., the polyharmonic spline function [32]. The main result of this paper is stated next.

Proposition 1 (Representer theorem)**.** *Let $\mathscr{P}_r$ and $\mathscr{P}_u$ denote the ground-truth distribution and distribution of unregularized reconstructions, respectively. Minimizers to the variational problem in Eq.* (3) *with $\lambda_{\mathrm{GP}} > 0$, and $p \in \mathbb{N}$ are of the form*

$$f^*_{p,n}(\mathbf{x}) = \frac{(-1)^{p+1}}{2\lambda_{\mathrm{GP}}} \left(\left(\mathscr{P}_r(\mathbf{x}) - \mathscr{P}_u(\mathbf{x})\right) * \phi_{p,n}(\mathbf{x})\right) + Q(\mathbf{x}), \tag{5}$$

where $Q(\mathbf{x})$ denotes a polynomial of degree $p-1$, and the polyharmonic spline kernel is defined as

$$\phi_{p,n}(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|^{2p-n}, & 2p-n < 0 \text{ or } n \text{ is odd}, \\ \|\mathbf{x}\|^{2p-n} \log(\|\mathbf{x}\|), & 2p-n \geq 0 \text{ and } n \text{ is even}. \end{cases}$$

The proof follows from [26]. The term Q denotes the homogenous component of the solution corresponding to the null space of Δ^p, and is intractable. Without compromising optimality, we set $Q(\mathbf{x}) = 0$ for the remainder of the discussion.

Fig. 2 shows the evaluation of the optimal regularization functionals obtained using the BSD500 dataset [33] evaluated around a point $\bar{\mathbf{x}}_r \sim \mathscr{P}_r$ along a line, along with f_{AR}. The AR and AR-ELA formulations, both, give rise to data-driven regularizers, unlike model-based regularizers like wavelet regularization which is independent of $\bar{\mathbf{x}}_r \sim \mathscr{P}_r$. The desired property of regularization functionals is more pronounced in the proposed formulation, i.e., the value of $f^*_{p,n}$ is low in the neighbourhood of $\bar{\mathbf{x}}_r \sim \mathscr{P}_r$, as compared to outside the neighbourhood, whilst the f_{AR} achieves lower values outside the neighbourhood of $\bar{\mathbf{x}}_r \sim \mathscr{P}_r$. We conjecture that f_{AR} would better approximate $f^*_{p,n}$ given more training data. The data-driven regularization functionals do not resemble model-based

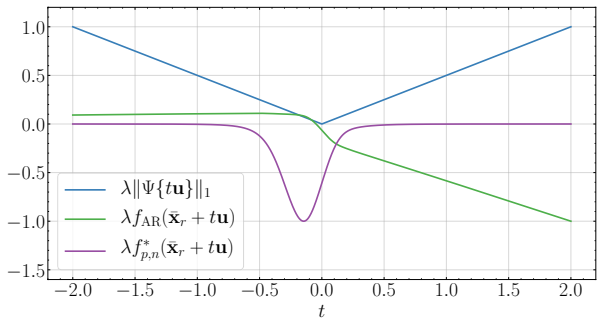


Fig. 2. Evaluation of regularization functionals on a line: wavelet regularizer, learnt regularizer f_{AR}, and $f^*_{p,n}$ with $2p-n=-3$. The direction $\mathbf{u}$ is the vector of all ones, and $\bar{\mathbf{x}}_r \sim \mathscr{P}_r$.

Algorithm 1: Solving inverse problems using AR-ELA

Input: Measurements $\mathbf{y}$, forward model $\mathbf{H}$, regularization parameters $\lambda, \lambda_{\mathrm{GP}} > 0$, training set $\{\bar{\mathbf{x}}_r\}$ and $\{\bar{\mathbf{y}}_r\}$

Compute: $\bar{\mathbf{x}}_u = \mathbf{H}^\dagger \mathbf{y}_r, \ \forall \mathbf{y}_r$ in training set

Initialization: $\mathbf{x}_0 = \mathbf{H}^\dagger \mathbf{y}$

for $k = 1, 2, \ldots$ *until convergence* **do**

  Choose step-size $\eta_k > 0$

  $\mathbf{x}_{k+1} = \mathbf{x}_k - \eta_k \left(\mathbf{H}^\top(\mathbf{H}\mathbf{x}_k - \mathbf{y}) + \lambda \nabla f^*_{p,n}(\mathbf{x}_k)\right)$

Output: $\mathbf{x}_{k+1}$

Algorithm 2: Solving inverse problems using AR-ELA+

Input: Measurements $\mathbf{y}$, forward model $\mathbf{H}$, regularization parameters $\lambda, \lambda_{\mathrm{GP}} > 0$, training set $\{\bar{\mathbf{x}}_r\}$ and $\{\bar{\mathbf{y}}_r\}$, momentum parameter $\alpha > 0$

Compute: $\bar{\mathbf{x}}_u = \mathbf{H}^\dagger \mathbf{y}_r, \ \forall \mathbf{y}_r$ in training set

Initialization: $\mathbf{x}_0 = \mathbf{z}_0 = \mathbf{H}^\dagger \mathbf{y}$

for $k = 1, 2, \ldots$ *until convergence* **do**

  Choose step-size $\eta_k > 0$

  $\mathbf{x}_{k+1} = \mathbf{z}_k - \eta_k \left(\mathbf{H}^\top(\mathbf{H}\mathbf{z}_k - \mathbf{y}) + \lambda \nabla f^*_{p,n}(\mathbf{z}_k)\right)$

  $\mathbf{z}_{k+1} = \mathbf{x}_{k+1} + \frac{k}{k+\alpha}(\mathbf{x}_{k+1} - \mathbf{x}_k)$

Output: $\mathbf{x}_{k+1}$

functions such as total-variation or wavelet regularization. In particular, the learnt regularization functionals are smooth, i.e., continuously differentiable, nonconvex, and depend on data $\mathbf{x}_r \sim \mathscr{P}_r$ and $\mathbf{x}_u \sim \mathscr{P}_u$. The corresponding iterative techniques for solving inverse problems are typically devoid of theoretical convergence guarantees.

3. SOLVING THE INVERSE PROBLEM USING GRADIENT DESCENT

Using the optimal regularization function in Proposition 1, the inverse problem requires solving the optmization problem:

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda f^*_{p,n}(\mathbf{x}).$$

Since $f^*_{p,n}$ is a continuously-differentiable function, the minimizing sequence using gradient-descent is given as:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta_k \left(\mathbf{H}^\top(\mathbf{H}\mathbf{x}_k - \mathbf{y}) + \lambda \nabla f^*_{p,n}(\mathbf{x}_k)\right), k \in \mathbb{N},$$

where $\eta_k > 0$ is the step-size chosen appropriately. The gradient of $f^*_{p,n}$ can be computed using sample estimates as

$$\nabla f^*_{p,n}(\mathbf{x}) = \frac{(-1)^{p+1}}{2\lambda_{\mathrm{GP}} N} \left[\sum_{\bar{\mathbf{x}}_r \sim \mathscr{P}_r} \nabla \phi_{p,n}(\mathbf{x} - \bar{\mathbf{x}}_r) - \sum_{\bar{\mathbf{x}}_u \sim \mathscr{P}_u} \nabla \phi_{p,n}(\mathbf{x} - \bar{\mathbf{x}}_u)\right],$$

where N is the number of training samples. Owing to the usage of Euler-Lagrange analysis in the AR framework, we call this method "AR-ELA," which is summarized in Algorithm 1.

Incorporating momentum: The natural extension of Algorithm 1 is to incorporate Nesterov's momentum [34] in the iterates to accelerate convergence. With the initialization $\mathbf{z}_0 = \mathbf{x}_0$, the

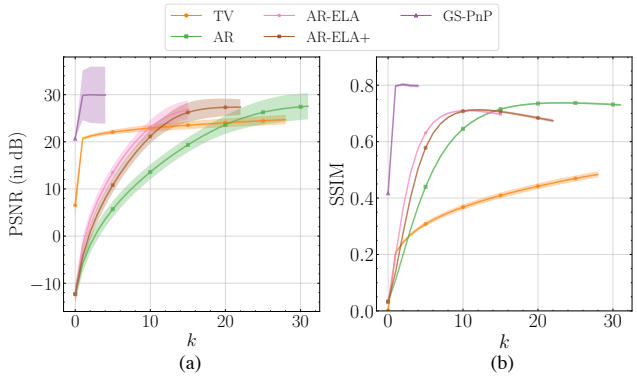


Fig. 3. Convergence analysis: (a) PSNR (in dB), (b) SSIM are shown as a function of the iteration index k. The proposed techniques AR-ELA and AR-ELA+ are similar in performance to AR, and outperform TV. AR-ELA+ converges the fastest.

minimizing sequence is given by

$$\mathbf{x}_{k+1} = \mathbf{z}_k - \eta_k \left(\mathbf{H}^\top (\mathbf{H}\mathbf{z}_k - \mathbf{y}) + \lambda \nabla f^*_{p,n}(\mathbf{z}_k) \right),$$
$$\mathbf{z}_{k+1} = \mathbf{x}_{k+1} + \frac{k}{k+\alpha}(\mathbf{x}_{k+1} - \mathbf{x}_k),$$

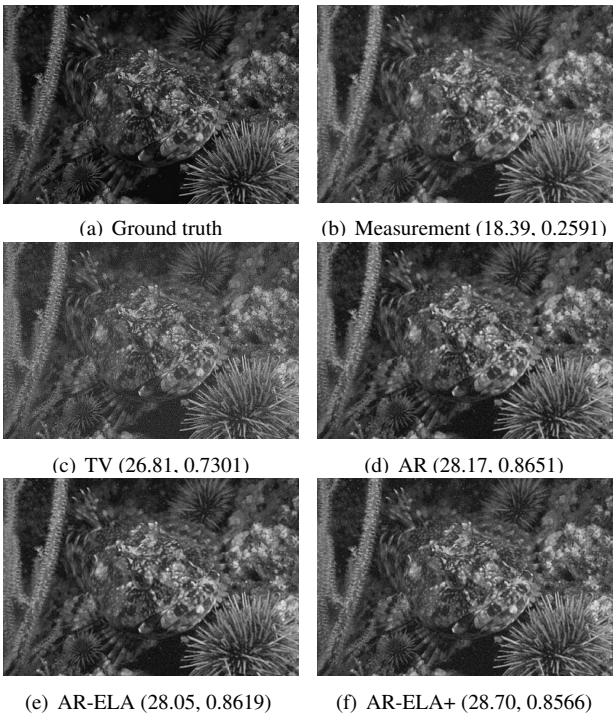
where $\alpha > 0$ is the *momentum parameter*. Owing to incorporating momentum in AR-ELA, we refer to this method as "AR-ELA+." A summary is provided in Algorithm 2.

4. EXPERIMENTAL RESULTS

We demonstrate our technique for image deconvolution on the BSD500 dataset [33], and compare the technique against AR [18] and total-variation (TV) minimization [7] and GS-PnP [35], which is the state-of-the-art supervised convergent plug-and-play technique. We use PSNR and SSIM as objective metrics. In image deconvolution, linear measurements of a ground-truth image $\mathbf{x}^*$ are obtained as $\mathbf{y} = \mathbf{h} * \mathbf{x}^* + \mathbf{w}$, where $\mathbf{h}$ denotes the convolution kernel, and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ denotes additive white Gaussian noise with standard deviation $\sigma > 0$. Throughout the experiments, we set the convolution kernel to a 5×5 Gaussian blur kernel with a standard deviation of 1, and set the noise level such that the SNR is 20 dB. We use the Wiener filter as the pseudoinverse operator to obtain unregularized reconstructions from the measurements. We use 300 images from the BSD500 dataset [33] for training and the remaining 200 images for testing.

In AR, the regularization functional is a deep convolutional neural network with an architecture similar to [18]. The network is trained with the loss function in Eq. (2), and Adam optimizer [36] with learning rate 10^{-3} and momentum parameters of $0.5, 0.9$ for 100 epochs.

We use Algorithms 1 and 2 with a constant step-size of $\eta_k = 10^{-1}$, $\alpha = 2$, $\lambda = 10^{-1}$ and $\lambda_{\text{GP}} = 10$. We set a tolerance of 4.5×10^{-3} on $\frac{\|\mathbf{x}_k - \mathbf{x}_{k-1}\|_2}{\|\mathbf{x}_k\|_2}$ or a maximum of 50 iterations, whichever occurs first, as the stopping criterion. Fig. 3 shows the average PSNR and SSIM over the test set as a function of the iteration index. We observe that the proposed techniques empirically converge in about 30 iterations with a similar performance to AR, and superior performance to TV minimization. AR-ELA+ converges faster as compared to AR-ELA and AR. Table 4 shows the average PSNR and SSIM with one standard deviation obtained after convergence.



(a) Ground truth (b) Measurement (18.39, 0.2591)
(c) TV (26.81, 0.7301) (d) AR (28.17, 0.8651)
(e) AR-ELA (28.05, 0.8619) (f) AR-ELA+ (28.70, 0.8566)

Fig. 4. Example reconstructions using the proposed techniques and benchmarks. The numbers indicate PSNR (in dB) and SSIM.

Table 1. Performance of the proposed techniques and benchmarks evaluated using PSNR↑ (in dB) and SSIM↑ on BSD500.

	TV [7]	AR [18]	AR-ELA	AR-ELA+	GS-PnP [35]
PSNR	24.71 ± 1.00	27.60 ± 2.66	26.98 ± 1.88	27.38 ± 1.63	29.89 ± 6.07
SSIM	0.48 ± 0.01	0.73 ± 0.00	0.69 ± 0.00	0.67 ± 0.00	0.80 ± 0.00

Fig. 4 shows example reconstructions using the proposed techniques and the benchmark methods. The supervised techniques have a superior performance in terms of PSNR and SSIM. The performance of AR-ELA and AR-ELA+ in terms of PSNR and SSIM, and in terms of visual assessment, are similar to that of AR.

5. CONCLUSIONS

We analyzed the solutions to the training problem in the AR framework using Euler-Lagrange analysis and showed that the optimal regularization function can be represented using the family of polyharmonic spline functions, and can be computed in closed form. The optimal regularization functional is continuously differentiable. Using this result, we solved the optimization problem using gradient descent (AR-ELA) and gradient descent with Nesterov's momentum (AR-ELA+). We demonstrated the technique to perform image deconvolution, and showed that it performs on par with the learning-based counterparts. The key feature of the proposed techniques is that they are data-driven approaches that do not require a rigorous prior training, unlike in AR where the regularization functional is parametrized using a neural network and heavily optimized using training data.

6. REFERENCES

[1] A. Ribes and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 84–99, 2008.

[2] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[3] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*. Springer, 2009.

[4] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics, 2011.

[5] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," *Russ. Acad. Sci. Dokl. Akad. Nauk*, vol. 151, no. 3, pp. 501–504, 1963.

[6] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

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

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

[9] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, 2017.

[10] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: Beyond analytical methods," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 20–36, 2018.

[11] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.

[12] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, "Learning maximally monotone operators for image recovery," *SIAM J. Imag. Sci.*, vol. 14, no. 3, pp. 1206–1237, 2021.

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

[14] 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 Process. Mag.*, vol. 37, no. 1, pp. 105–116, 2020.

[15] R. G. Gavaskar, C. D. Athalye, and K. N. Chaudhury, "On plug-and-play regularization using linear denoisers," *IEEE Trans. Image Process.*, vol. 30, pp. 4802–4813, 2021.

[16] J. Scarlett, R. Heckel, M. R. D. Rodrigues, P. Hand, and Y. C. Eldar, "Theoretical perspectives on deep learning methods in inverse problems," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, pp. 433–453, 2022.

[17] S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb, "Learned reconstruction methods with convergence guarantees: A survey of concepts and applications," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 164–182, 2023.

[18] S. Lunz, O. Öktem, and C.-B. Schönlieb, "Adversarial regularizers in inverse problems," in *Proc. Adv. Neural Inf. Process. Sys.*, 2018.

[19] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb, "Learned convex regularizers for inverse problems," *arXiv preprint arXiv:2008.02839*, 2020.

[20] C. Villani *et al.*, *Optimal transport: Old and new*, vol. 338. Springer, 2009.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[23] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, 2018.

[24] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 214–223, 2017.

[25] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Sys.*, 2017.

[26] S. Asokan and C. S. Seelamantula, "Data interpolants—That's what discriminators in higher-order gradient-regularized GANs are," *arXiv preprint arXiv:2306.00785*, 2023.

[27] S. Asokan and C. S. Seelamantula, "Euler-Lagrange analysis of generative adversarial networks," *J. Mach. Learn. Res.*, vol. 24, no. 126, pp. 1–100, 2023.

[28] I. M. Gelfand and R. A. Silverman, *Calculus of Variations*. Courier Corporation, 2000.

[29] T. Kurokawa, "Riesz potentials, higher Riesz transforms and Beppo-Levi spaces," *Hiroshima Math. J.*, vol. 18, no. 3, pp. 541–597, 1988.

[30] M. Arigovindan, M. Suhling, P. Hunziker, and M. Unser, "Variational image reconstruction from arbitrarily spaced samples: A fast multiresolution spline solution," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 450–460, 2005.

[31] J. Adler and S. Lunz, "Banach Wasserstein GAN," in *Proc. Adv. Neural Inf. Process. Sys.*, 2018.

[32] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*, pp. 85–100, Springer, 1977.

[33] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011.

[34] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $\mathcal{O}(\frac{1}{k^2})$," *Russ. Acad. Sci. Dokl. Akad. Nauk*, vol. 269, no. 3, pp. 543–547, 1983.

[35] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent plug-and-play," in *Proc. Int. Conf. Learn. Rep. (ICLR)*, 2022.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.