Test-Time Forward Model Adaptation for Seismic Deconvolution

Peimeng Guan, Naveed Iqbal, *Senior Member, IEEE*, Mark A. Davenport, *Senior Member, IEEE*, and Mudassir Masood, *Senior Member, IEEE*

***Abstract*—Seismic deconvolution is essential for extracting layer information from noisy seismic data, but it is an ill-posed problem with nonunique solutions. Inspired by classical optimization approaches, model-based deep learning architectures, such as loop unrolling (LU) methods, unfold the optimization process into iterative steps and learn gradient updates from data. These architectures rely on well-defined forward models, but in real seismic deconvolution scenarios, these models are often inaccurate or unknown. Previous approaches have addressed model uncertainty by training robust networks, either passively or actively. However, these methods require a large number of adversarial examples and diverse data structures, often necessitating retraining for unseen forward model structures, which is resource-intensive. In contrast, we propose a more efficient test-time adaptation (TTA) method for the LU architecture, which refines the forward model during inference. This approach incorporates physical principles into the reconstruction process, enabling higher quality results without the need for costly retraining. The code is available at: https://github.com/InvProbs/A-adaptive-seis-deconv**

***Index Terms*—Deep learning, loop unrolling (LU), model mismatch, seismic deconvolution, test-time adaptation (TTA).**

I. Introduction

ESTIMATING subsurface reflectivity from seismic traces is a fundamental task in seismic data processing. During a geophysical survey, a source wave generated by a vibroseis truck propagates through the Earth and reflects at layer boundaries, with the reflected signals recorded by geophones. These recordings, known as *seismic traces*, result from the convolution of the source wavelet with a sparse reflectivity series that marks impedance contrasts between layers [1]. The sparsity pattern depends on the subsurface structure and is generally unknown.

Seismic deconvolution aims to recover the reflectivity sequence from observed traces by inverting this convolution process [1], [2]. This inverse problem is ill-posed and nonunique due to measurement noise and the band-limited nature of the wavelet, which suppresses high-frequency components and blurs closely spaced reflectors [3], [4]. Formally, the problem is modeled as

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

where $\boldsymbol{y} \in \mathbb{R}^n$ is the seismic trace, $\boldsymbol{x} \in \mathbb{R}^n$ is the reflectivity sequence, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is the convolution matrix derived from the source wavelet, and $\boldsymbol{\epsilon}$ is the unknown additive noise.

When the forward model is precisely known, classical optimization approaches like ISTA [5] and FISTA [6] solve deconvolution using iterative optimization with sparsity-promoting regularization. However, these methods are slow and require careful parameter tuning. Data-driven methods [7], [8], [9], [10], including U-Net and sparse-promoting networks [11] directly learn mapping from traces to reflectivities. Although efficient, these models ignore the physical forward model, limiting their robustness and generalization.

Model-based deep learning methods, such as *loop unrolling* (LU) architectures [12], unfold iterative optimization into neural network blocks, explicitly incorporating the forward model into the recovery process [12], [13]. This results in interpretable and high-quality reconstructions across tasks like medical imaging, image deblurring, and compressive sensing [14], [15], [16], [17], [18], [19], [20], [21], [22]. Recently, such architectures have gained traction in seismic inversion. For example, [23] employs ℓ_1-based minimax-convex regularization, while other works leverage gradient-based solvers [24], [25], projected gradient descent [26], and augmented Lagrangian methods [27]. Untrained unrolled networks have also been used as deep priors [28]. However, these methods require accurate knowledge of the forward model during training and inference, relying on supervised datasets of ($\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{A}$) tuples. As shown in [29], their performance degrades significantly when the forward model is inaccurate.

In seismic reconstruction, test-time wavelet estimation errors can introduce forward model inaccuracies, limiting real-world performance despite precise models used during training. Existing methods improve robustness by training LU architectures with adversarial perturbations [30] or learning model mismatches [29], but these require costly retraining with carefully designed errors. To address this, we propose a test-time forward model adaptation for LU architectures trained on the correct forward model, enhancing reconstruction without additional training.

Test-time adaptation (TTA) is a learning paradigm that aims to improve a model's performance at inference time by adapting it to the test data distribution, without requiring access to labeled examples from the target domain during training. TTA is particularly valuable in scenarios where models are

Received 23 April 2025; revised 9 July 2025; accepted 4 August 2025. Date of publication 12 August 2025; date of current version 21 August 2025. This work was supported in part by the Center for Energy and Geo Processing (CeGP) and in part by the Deanship of Research Oversight and Coordination (DROC), King Fahd University of Petroleum and Minerals (KFUPM) under Project GTEC2013. *(Corresponding author: Mudassir Masood.)*

Peimeng Guan and Mark A. Davenport are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: pguan6@gatech.edu).

Naveed Iqbal and Mudassir Masood are with the Department of Electrical Engineering and the Interdisciplinary Research Center for Communication Systems and Sensing (IRC-CSS), King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia (e-mail: mudassir@kfupm.edu.sa).

Digital Object Identifier 10.1109/LGRS.2025.3598143

subject to distribution shifts at deployment [31]. By exploiting the structure of incoming test data, typically through self-supervised objectives [32] and entropy-based criteria such as entropy minimization [33], [34], TTA enables models to maintain or even improve performance under out-of-distribution conditions, without retraining or manual annotation. In the context of seismic deconvolution, our proposed TTA framework offers a promising strategy for enhancing robustness against forward model mismatches that commonly arise in real-world applications.

A. Our Contributions

This work proposes a TTA method to address the forward model, leveraging a pretrained model-based architecture that is trained using synthetic data with a precisely known forward model. The proposed approach has the following advantages.

1) Robust to Real-World Data Variability: Synthetic training data simulates ideal conditions, but real-world data often deviates. An adaptive approach enables the model to handle this variability effectively.

2) Cost-Effective: Retraining models with new data is costly. TTA allows the model to handle new conditions without the expense of retraining.

II. Problem Description

An estimate of the reflectivity sequence $\hat{\boldsymbol{x}}$ can be obtained by solving the following optimization problem:

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \gamma\, r(\boldsymbol{x}). \tag{2}$$

Here, $\|\boldsymbol{y}-A\boldsymbol{x}\|_2^2$ penalizes the data misfit using current estimate of $\boldsymbol{x}$. Furthermore, $r : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a regularization function. The choice of the r depends on the prior beliefs of the underlying signal $\boldsymbol{x}$ and the computational feasibility. For example, ℓ_2-norm encourages minimum norm solutions, while ℓ_0-norm encourages sparse solutions [35]. In addition, γ is the regularization hyper-parameter that is usually well-tuned to balance the data misfit and the regularization term.

The problem in (2) can be solved via iterative optimization methods, such as proximal gradient descent, where in each iteration, the proximal operator of r is applied to the gradient update of the data misfit term, as shown below

$$\hat{\boldsymbol{x}}_{k+1} = \text{prox}_{\eta,r}(\hat{\boldsymbol{x}}_k + \eta \boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}_k)). \tag{3}$$

The proximal operator enforces the structure that the regularization function r attempts to encourage. For simple regularizers like ℓ_1 and ℓ_2, the proximal operators have closed-form solutions but for general regularizers, closed-form solutions may not exist. To overcome these challenges and accelerate optimization, learning-based methods like LU mimic proximal gradient iterations with trainable parameters, enabling data-driven, efficient updates.

III. Methodology

This section reviews related methods and introduces the proposed LU with test-time forward model adaptation method, denote as LU-TTA. Trained with an accurate A, LU-TTA adapts to inaccuracies during real-world inference, ensuring robust performance.

A. LU Method

This method [12], [36], [37], [38] unfolds the optimization problem in (2) into a sequence of iterative steps, with the regularization updates (i.e., *prox*) replaced by shared-weights neural networks. This approach leverages data-driven learning to achieve more accurate and robust reconstruction results. Training the proximal LU starts with initializing the input $\hat{\boldsymbol{x}}_0 = y$ and a fixed number of iterations K. For each iteration $k = 1, \ldots, K$, the network updates the estimate by computing $\hat{\boldsymbol{x}}_k + \eta \boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}_k)$, then passed through the learnable proximal operator to obtain the next refined reconstruction. After K iterations, the ultimate output $\hat{\boldsymbol{x}}_K$ is compared to the ground-truth x for training.

B. A-Adaptive LU Method

At inference time, LU requires an accurate forward model A to avoid artifacts [14], [29], which may not always be available in practice. A recent work [29] proposes to use an untrained neural network to approximate the forward model mismatch for each data instance along with reconstruction. Assuming the initial forward model A_0 is the Toeplitz matrix formed using the approximated wavelet ω_0. The forward model residual network $f_\theta : x \mapsto y$ aims to fix the error due to model mismatch

$$y = A_0 x + f_\theta(x) + \epsilon.$$

Guan et al. [29] proposed to introduce an auxiliary variable z, and solve the following optimization instead:

$$\min_{x,z,\theta} \frac{1}{2}\|y - A_0 z - f_\theta(z)\|_2^2 + \gamma r(x) + \tau\|f_\theta(z)\|_2^2 + \lambda\|x - z\|_2^2. \tag{4}$$

The third term ensures that the forward model residual is small, and the last term regulates the difference between the reconstruction x and the auxiliary variable z. The variables can be solved iteratively with step size η for time steps $k = 1, 2, \ldots, K$

$$\begin{aligned}
\hat{z}_{k+1} &= \arg\min_{z} \frac{1}{2}\|y - A_0 z - f_{\theta_k}(z)\|_2^2 + \tau\|f_{\theta_k}(z)\|_2^2 \\
&\quad + \lambda\|\hat{x}_k - z\|_2^2, \\
\theta_{k+1} &= \arg\min_{\theta} \frac{1}{2}\|y - A_0\hat{z}_{k+1} - f_\theta(\hat{z}_{k+1})\|_2^2 + \tau\|f_\theta(\hat{z}_{k+1})\|_2^2, \\
\hat{x}_{k+1} &= \text{prox}_r(\hat{x}_k - \eta(\hat{x}_k - \hat{z}_{k+1})).
\end{aligned} \tag{5}$$

Notice that f_θ is an untrained network whose weights θ are updated for each instance along with the reconstruction. The proximal operator, prox_r, is replaced by a neural network reffered to as the proximal network, which takes the input $\hat{x}_k - \eta(\hat{x}_k - \hat{z}_{k+1})$ at each iteration. After the final iteration K, the reconstruction $\hat{x}_K$ is compared to the ground-truth x, and only the weights in the proximal network are optimized using backpropagation.

The method proposed in [29] generates inaccurate forward models during training to teach the network to correct errors in the forward model at inference. In contrast, our work builds on this adaptation approach, demonstrating that an LU network trained with accurate forward models can still adapt to inaccuracies during inference. This strategy effectively enhances the network's generalization by incorporating forward model adaptation solely at the evaluation stage.

C. Proposed Method—LU-TTA

In this work, instead of training LU with deliberately corrupted forward models, we propose to train the LU in a standard way with accurate A, and introduce forward model correction only during inference to reduce training costs. This significantly reduces training complexity while preserving robustness. When training LU, the gradient of the least-squares term in (2) incorporates the forward model and indicates the update direction to match the noisy trace, and the proximal network in (2) is a correction term that learns complicated structures of x. Thus, the learned proximal network can also be viewed as a task-specific denoising process. Over K iterations, the same proximal operator is essentially trained to denoise the reflectivity sequence at different noise levels (larger noise level with smaller k) and noise patterns (more structured noise with smaller k). Thus, prior works [29] have observed that model-based architectures such as LU are more robust and can generalize better than learning a direct reconstruction network.

Algorithm 1 LU With TTA (LU-TTA)

```
Load the pre-trained proximal network trained from LU
for each batch of evaluation data (y, A_0) do
    Initialize x̂_0 = ẑ_0 = A_0^⊤ y
    Initialize the forward model residual network f_θ
    for k = 1,...,K do
        Update ẑ_k, θ_k, and x̂_k, according to (5)
    end for
end for
```

To mitigate the effects of forward model inaccuracies without retraining, the proposed LU-TTA introduces a forward model correction function f_θ that refines the estimated forward model during inference. Given the current reconstruction estimate, f_θ aims to minimize the second objective function in (5), where $\|y - A_0\hat{z}_{k+1} - f_\theta(\hat{z}_{k+1})\|_2^2$ ensures that the corrected forward model aligns with the observed trace, while $\|f_\theta(\hat{z}_{k+1})\|_2^2$ prevents excessive modifications to the original forward model. During inference, the correction function f_θ adapts dynamically for each data instance, enabling LU-TTA to handle forward model errors without requiring extensive retraining. Algorithm 1 outline the LU-TTA.

The ability of LU-TTA to adapt to forward model inaccuracies during inference stems from the complementary roles of the learned proximal operator and the forward model correction function. The proximal network, trained on ideal forward models, acts as a task-specific denoiser that refines the reflectivity estimate at each iteration. Since LU implicitly learns a structured optimization process, the trained proximal network generalizes well across different noise conditions. This allows LU-TTA to improve reflectivity sequence estimation even when faced with unseen physical conditions.

IV. Experiments and Discussions

In the experiments, we compare the proposed evaluation scheme with 1) a U-Net that directly learns the deconvolution process in a supervised manner without utilizing the forward model and 2) an LU network with DnCNN as the proximal operator that is trained using the correct forward model. The U-Net serves as a baseline where the forward model A does not explicitly appear in the reconstruction, while the LU network assumes that A is precisely known during training. We first validate the proposed approach using synthetic data with controllable A and further demonstrate its effectiveness on real seismic data without ground truth. The performance of the proposed method is evaluated using three different metrics when the ground-truth data is available: mean-squared error (mse), correlation coefficient, and reconstruction quality Q. These metrics measure the discrepancy between the true and estimated reflectivities.

A. Training on Synthetic Dataset

All networks are trained on the same set of synthetic data, generated according to the method outlined in [39]. The 2-D data comprises $m = 352$ traces per shot, each collecting $n = 352$ data points per trace along the Earth's depth, sampled at 500 Hz. A 40-Hz Ricker wavelet is employed to generate the observed trace $\boldsymbol{y}$. Additive white Gaussian noise is introduced to the data, with a variance of 0.1. Traces are normalized without mean subtracted to preserve zero magnitude in reflectivity. The same procedure applies during inference. While U-Net training is independent of the forward model A, LU training assumes the correct A as used in data generation.

B. Evaluation With Known A

While the main goal of this article is to address inaccuracies in the forward model during evaluation, we first present baseline results on synthetic data from the training distribution, assuming a known forward model. We also extend the evaluation to the Marmousi2 model [40], a widely used benchmark for complex seismic processing. This model simulates a real seismic deconvolution problem, where the impedance is derived from the density-velocity product, and the true reflectivity is computed from vertical impedance changes. The observed traces are generated by convolving a 40-Hz Ricker wavelet with the reflectivity profile. The first two columns of Table I show the evaluation results, where LU outperforms U-Net by leveraging A in iterative refinement.

C. Evaluation on Marmousi2 Model With Inaccurate A

We simulate the multiplicative error in A using the Marmousi2 model. Multiplicative errors are frequently encountered in seismic deconvolution, primarily due to uncertainties associated with source wavelets, attenuation effects, and other environmental factors [41]. We perturb the forward model by introducing zero-mean random Gaussian noise, applying both low variance (0.05) and high variance (0.1) levels to assess the impact of noise on the model's accuracy. Furthermore, small noises are added to the measurement domain to simulate additive noises in seismic exploration problems. Note that both the multiplicative and additive noise are treated as unknown during the evaluation, mimicking real-world scenarios where such errors cannot be precisely quantified. The two rightmost columns in Table I compare the Marmousi2 evaluation with inaccurate A. Upon analysis,

TABLE I

AVERAGE TESTING MSE, γ, AND Q FOR THE SYNTHETIC DATA WITH ACCURATE A (COLUMN 2), MARMOUSI2 DATASET WITH ACCURATE A (COLUMN 3), MARMOUSI2 DATASET WITH LOW NOSE LEVEL IN A (COLUMN 4), AND MARMOUSI2 DATASET WITH HIGH NOISE LEVEL IN A (COLUMN 5) USING DIFFERENT METHODS, WHERE THE BEST PERFORMANCES FOR EACH METRIC ARE IN BOLD

Models	2D Synthetic (known $\boldsymbol{A}$)	Marmousi2 (known $\boldsymbol{A}$)	Marmousi2 (low error in $\boldsymbol{A}$)	Marmousi2 (high error in $\boldsymbol{A}$)
U-Net	0.001744 / 0.632 / 2.362	0.003162 / 0.721 / 3.543	0.009370 / 0.263 / 0.313	0.009180 / 0.072 / 0.024
LU	**0.000991 / 0.815 / 5.110**	**0.002145 / 0.877 / 7.234**	0.008087 / 0.418 / 0.837	0.006193 / 0.264 / 0.315
LU-TTA	-	-	**0.003434 / 0.610 / 2.023**	**0.003976 / 0.482 / 1.660**

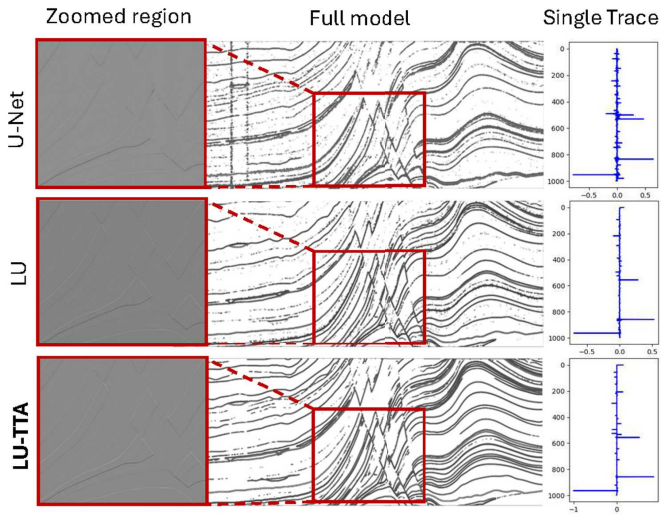

Fig. 1. Reconstruction results on Marmousi2 synthetic data. From left to right show a zoomed-in view region, the full model, and a single trace reconstruction comparison across different methods.

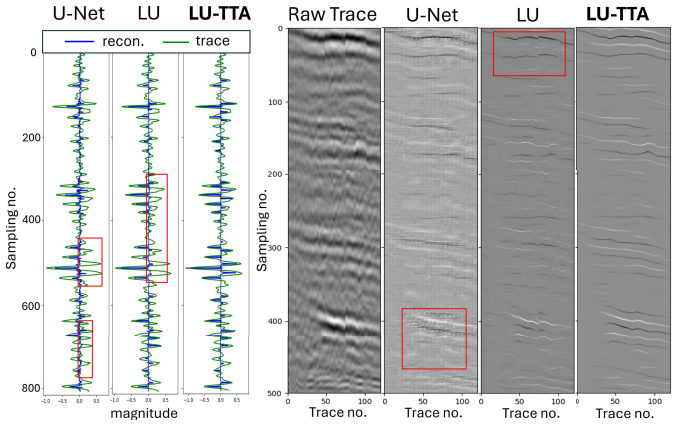

Fig. 2. Reconstruction results on real seismic data in Texas. The left shows the 1-D trace at trace no. 40 (green) with the reconstructed reflectivity (blue) for each method, while the right presents the 2-D trace and corresponding reconstructed reflectivity sequences. Notable distortions are marked with red boxes.

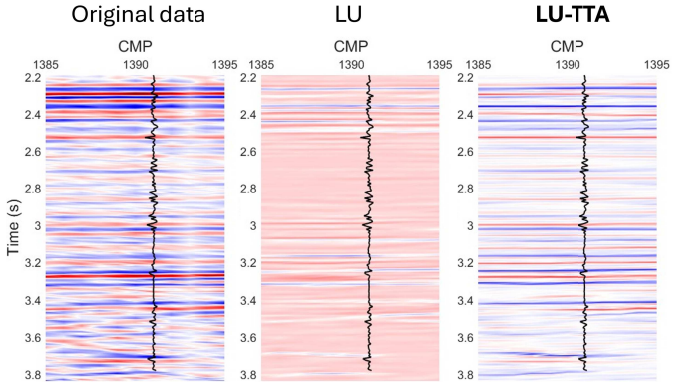

Fig. 3. Reconstruction results on real seismic data in Alaska. The well-log data (in black) are shown alongside the reconstructions.

it becomes evident that although the LU method exhibits greater robustness to multiplicative noise compared to U-Net. Conversely, LU-TTA demonstrates a significant improvement in handling these errors. This technique effectively mitigates the adverse effects of multiplicative noise, resulting in cleaner and more accurate reconstructions of the seismic data as shown in Fig. 1.

D. Evaluation on Real Seismic Data

Finally, we evaluate the proposed method on two real-world datasets: a 2-D landline from East Texas, USA [1], and a line from Alaska, USA. The Texas dataset contains 18 shots and 594 traces per shot, with each trace consisting of 1501 samples sampled at 500 Hz. A common midpoint (CMP) gather is extracted from the raw data, segmented into overlapping patches, and zero-padded to fit the network input. Missing traces are muted (zero-filled). The Alaska dataset is sourced from Line 7X-75 of the National Petroleum Reserve–Alaska (NPRA) Legacy Data Archive, provided by the USGS (1976), and includes well-log data for reference.

These real datasets serve as practical benchmarks for evaluating the robustness of LU-TTA under realistic seismic deconvolution scenarios. Unlike synthetic data with known forward models, real seismic data involves various uncertainties. The true source wavelet is unknown and may include multiplicative noise in A, structural variations, and attenuation effects. Additional sources of forward model error include subsurface heterogeneity, wavelet estimation inaccuracies, sensor misalignment, acquisition noise, and near-surface scattering. For evaluation, we estimate the source wavelet using a 40-Hz Ricker wavelet for both LU and LU-TTA. Fig. 2 shows the observed traces from Texas, and reconstructions from U-Net, standard LU, and LU-TTA. The U-Net result, which does not incorporate the forward model, suffers from a noticeable quality drop due to distribution shift, making true reflectivity hard to distinguish from artifacts. Standard LU, while leveraging A, cannot adapt to its inaccuracies, leading to residual artifacts. In contrast, LU-TTA dynamically refines the forward model at inference, producing cleaner, sparser reconstructions that better capture geological structures. Fig. 3 shows that

LU-TTA produces reconstructions on the Alaska dataset that align more closely with the well-log data than LU alone. It delivers more accurate and geologically plausible results. Overall, LU-TTA demonstrates superior adaptability to real-world model discrepancies, underscoring its practical utility in seismic exploration.

V. Conclusion

We propose LU-TTA to address forward model mismatch at inference for seismic deconvolution. By combining a pretrained LU with test-time forward model adaptation, LU-TTA maintains the efficiency and interpretability of model-based learning while improving robustness. Unlike traditional methods requiring inaccurate forward models during training, LU-TTA is trained with accurate models and adapts dynamically during evaluation, reducing training costs and computational overhead. Experiments show that LU-TTA outperforms conventional LU and deep learning methods on synthetic benchmarks, the Marmousi2 model, and real seismic data, delivering reliable reconstructions under unknown forward models.

References

[1] W. A. Mousa and A. A. Al-Shuhail, "Processing of seismic reflection data using MATLAB," *Synth. lectures signal Process.*, vol. 5, no. 1, pp. 1–97, 2011.

[2] D. W. Oldenburg, T. Scheuer, and S. Levy, "Recovery of the acoustic impedance from reflection seismograms," *Geophysics*, vol. 48, no. 10, pp. 1318–1337, Oct. 1983.

[3] M. van der Baan and D.-T. Pham, "Robust wavelet estimation and blind deconvolution of noisy surface seismics," *Geophysics*, vol. 73, no. 5, pp. V37–V46, Sep. 2008.

[4] M. G. Bostock and M. D. Sacchi, "Deconvolution of teleseismic recordings for mantle structure," *Geophys. J. Int.*, vol. 129, no. 1, pp. 143–152, Apr. 1997.

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

[6] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[7] B. Russell, "Machine learning and geophysical inversion—A numerical study," *Lead. Edge*, vol. 38, no. 7, pp. 512–519, Jul. 2019.

[8] Z. Gao et al., "A deep-learning-based generalized convolutional model for seismic data and its application in seismic deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4503117.

[9] D. Pereg, I. Cohen, and A. A. Vassiliou, "Sparse seismic deconvolution via recurrent neural network," *J. Appl. Geophysics*, vol. 175, Apr. 2020, Art. no. 103979.

[10] S. Phan and M. K. Sen, "Seismic nonstationary deconvolution with physics-guided autoencoder," in *Proc. 1st Int. Meeting Appl. Geosci. Energy Expanded Abstr.*, Sep. 2021, pp. 1635–1640.

[11] X. Chai et al., "Deep learning for multitrace sparse-spike deconvolution," *Geophysics*, vol. 86, no. 3, pp. V207–V218, May 2021.

[12] 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, Mar. 2021.

[13] M. Andrychowicz et al., "Learning to learn by gradient descent by gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3988–3996.

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

[15] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep MRI reconstruction: Unrolled optimization algorithms meet neural networks," 2019, *arXiv:1907.11711*.

[16] P. Putzky and M. Welling, "Invert to learn to invert," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 446–456.

[17] Y. Li, M. Tofighi, J. Geng, V. Monga, and Y. C. Eldar, "Deep algorithm unrolling for blind image deblurring," 2019, *arXiv:1902.03493*.

[18] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A deep learning approach for image compressive sensing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 521–538, Mar. 2020.

[19] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya, "Dense recurrent neural networks for inverse problems: History-cognizant unrolling of optimization algorithms," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1280–1291, Oct. 2020.

[20] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[21] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 10–18.

[22] S. Lohit, D. Liu, H. Mansour, and P. T. Boufounos, "Unrolled projected gradient descent for multi-spectral image fusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7725–7729.

[23] S. Mache, P. Kumar Pokala, K. Rajendran, and C. Sekhar Seelamantula, "DuRIN: A deep-unfolded sparse seismic reflectivity inversion network," 2021, *arXiv:2104.04704*.

[24] K. Torres and M. D. Sacchi, "Least-squares reverse time migration via deep learning-based updating operators," *Geophysics*, vol. 87, no. 6, pp. S315–S333, 2022, doi: 10.1190/geo2021-0491.1.

[25] P. Sun, F. Yang, H. Liang, and J. Ma, "Full-waveform inversion using a learned regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5920715.

[26] H. Peng, I. Vasconcelos, and M. Ravasi, "Image-domain seismic inversion by deblurring with invertible recurrent inference machines," *Geophysics*, vol. 89, no. 2, pp. R121–R136, Mar. 2024.

[27] H. Zhou, T. Xie, Q. Liu, and S. Hu, "An augmented Lagrangian method-based deep iterative unrolling network for seismic full-waveform inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5914613.

[28] W. Zhang, M. Ravasi, J. Gao, and Y. Shi, "Deep-unrolling architecture for image-domain least-squares migration," *Geophysics*, vol. 89, no. 3, pp. S215–S234, May 2024, doi: 10.1190/geo2023-0428.1.

[29] P. Guan, N. Iqbal, M. A. Davenport, and M. Masood, "Solving inverse problems with model mismatch using untrained neural networks within model-based architectures," *Trans. Mach. Learn. Res.*, May 2024. [Online]. Available: https://openreview.net/forum?id=XHEhjDxPDl

[30] J. Hu, S. Shoushtari, Z. Zou, J. Liu, Z. Sun, and U. S. Kamilov, "Robustness of deep equilibrium architectures to changes in the measurement model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[31] J. Liang, R. He, and T. Tan, "A comprehensive survey on test-time adaptation under distribution shifts," *Int. J. Comput. Vis.*, vol. 133, no. 1, pp. 31–64, Jan. 2025.

[32] Y. Sun, X. Wang, Z. Liu, J. J. Miller, A. A. Efros, and M. Hardt, "Test-time training with self-supervision for generalization under distribution shifts," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 9229–9248.

[33] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 6028–6039.

[34] H. Yan, Y. Guo, and C. Yang, "Augmented self-labeling for source-free unsupervised domain adaptation," in *Proc. Workshop Distrib. Shifts, Connecting Methods Appl. (NeurIPS)*, Mar. 2021.

[35] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, 1979.

[36] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[37] H. Sreter and R. Giryes, "Learned convolutional sparse coding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2191–2195.

[38] J. Sulam, A. Aberdam, A. Beck, and M. Elad, "On multi-layer basis pursuit, efficient algorithms and convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 1968–1980, Aug. 2020.

[39] N. Iqbal, E. Liu, J. H. Mcclellan, and A. A. Al-Shuhail, "Sparse multichannel blind deconvolution of seismic data via spectral projected-gradient," *IEEE Access*, vol. 7, pp. 23740–23751, 2019.

[40] G. S. Martin, R. Wiley, and K. J. Marfurt, "Marmousi2: An elastic upgrade for marmousi," *Lead. Edge*, vol. 25, no. 2, pp. 156–166, Feb. 2006.

[41] A. Bakulin, D. Neklyudov, and I. Silvestrov, "Multiplicative random seismic noise caused by small-scale near-surface scattering and its transformation during stacking," *Geophysics*, vol. 87, no. 5, pp. V419–V435, Sep. 2022.