Formulating Full Waveform Inversion as a Gradient Flow Problem

Idris Oduola

July 14, 2025

1 Motivation

1.1 Full Waveform Inversion (FWI)

Full Waveform Inversion (FWI) is an advanced inverse-problem technique widely used in geophysics for accurately determining subsurface properties, primarily the velocity field $u(x)$. It achieves this by numerically matching simulated seismic waveforms to observed seismic data. The procedure involves a forward problem coupled with an inverse problem, iteratively.

1.1.1 Forward Problem

Given an initial velocity $u(x)$, FWI starts with a forward modeling step which involves solving the acoustic wave equation in a spatial domain $\Omega \subset R^d$:

$$\nabla^2 p(x,t) - \frac{1}{u(x)^2}\frac{\partial^2 p(x,t)}{\partial t^2} = s(x,t), \qquad x \in \Omega, t > 0, \tag{1}$$

where

- $p(x,t)$ is called a pressure wavefield at position x and time t.
- $u(x)$ is the velocity which is a function of x.
- $s(x,t)$ is a known seismic source term.

An appropriate boundary condition can be chosen, whereas the initial condition is typically:

$$p(x,0) = \frac{\partial p}{\partial t}(x,0) = 0, \qquad x \in \Omega. \tag{2}$$

1.1.2 Inverse Problem

The velocity $u(x)$ describes the speed with which the wave propagates through a certain medium (e.g, underground), and this enables seismic methods to produce high-resolution, structured images of the medium. Unfortunately, we do not have the velocity except some observed pressure wavefields $p_{obs}(x,t)$ at some finite locations in Ω. This leads to an inverse problem, for which FWI seeks to minimize a functional $\mathcal{J}(p(u))$ that measures the difference between the observed wavefield and some simulated wavefield $\hat{p}(x,t)$ obtained by solving equation 1 numerically.

$$\mathcal{J}(p(u)) = \frac{1}{2}\sum_r \int_0^T |\hat{p}(x_r,t;u) - p_{obs}(x_r,t)|^2 dt, \tag{3}$$

where,

- x_r are the locations in the domain of the observed wavefields,
- T is the total time.

The inverse method uses a gradient-based optimization method to obtain the velocity $u(x)$, given a trial velocity $u_0(x)$,

$$u_{k+1}(x) = u_k(x) - \alpha_k \nabla_u \mathcal{J}(p(u_k))(x), \tag{4}$$

where α_k is the step length and ∇_u is the derivative with respect to u.

1.1.3 Algorithm for FWI

- Initialize a trial $u_0(x)$.
- Solve wave equation to compute wavefields $\hat{p}(x, t; u_k)$.
- Evaluate the functional $\mathcal{J}(p(u_k))$.
- Calculate the gradient $\nabla_u \mathcal{J}(p(u_k))$.
- Update the velocity using equation (4).

1.1.4 Challenges of FWI

FWI faces several critical challenges in obtaining the velocity $u(x)$ that minimizes (4):

- **Nonlinearity and non-convexity:** The inversion process may converge to local minima, limiting the accuracy of reconstructed velocities.
- **Sensitivity to initial conditions:** The quality of the initial velocity u_0 greatly impacts convergence and accuracy.
- **Computational expense:** Solving PDEs in every iteration makes the inversion computationally intensive.
- **Cycle-skipping:** Incorrect phase alignment between predicted and observed waveforms may prevent convergence to the global optimum.

1.2 Machine Learning for FWI

Scientific machine learning (SciML) has emerged as a powerful approach in computational science and engineering, primarily driven by advances in deep learning [1]. These methods excel at capturing complex mappings between high-dimensional spaces, which makes them particularly suitable for addressing challenging computational problems such as those encountered in FWI. SciML methodologies can be broadly categorized into two primary frameworks: function-to-function mappings (e.g, Physics-Informed Neural Networks (PINNs) [2]) and operator mappings (e.g, neural operators [3, 4]).

Neural operators are especially compelling due to their capability to learn PDE solution operators, directly mapping input functions—such as initial conditions, boundary conditions, or forcing terms—to PDE solutions. Formally, these operators approximate a mapping $\mathcal{G}$ defined as:

$$\mathcal{G} : X \mapsto Y,$$

where X and Y are two infinite-dimensional Banach spaces of functions and $y = \mathcal{G}(v)$ corresponding to the PDE solution.

In FWI, traditional methods rely heavily on repeatedly solving partial differential equations (PDEs), which is a computationally expensive process. Machine learning techniques, however, offer promising alternatives by serving as surrogate models. Specifically, they can directly learn mappings between wavefields p and their corresponding velocity models u. Once trained, these surrogate models significantly reduce computational costs by predicting velocities rapidly, circumventing the repeated, costly numerical PDE solvers. Additionally, a critical challenge in traditional FWI is the sensitivity to the initial velocity model u_0. Surrogate models trained via machine learning can aid in addressing this issue by providing informed initial velocity models. Improved initial estimates facilitate faster convergence and enhance the overall efficiency of inversion procedures.

Several machine learning architectures have been applied in FWI, including Convolutional Neural Network (CNN) with an encoder-decoder structure [5], Generative Adversarial Network (GAN) with modified encoder-decoder structure [6], and multiple neural operator variants such as Fourier-DeepONet [7], Inversion-DeepONet [8], En-DeepONet [9], among others [10, 11].

Despite their advantages, current SciML approaches exhibit significant limitations. Primarily, they are characterized by their black-box nature, limiting physical interpretability. Additionally, their heavy reliance on training data restricts their generalizability to new geological scenarios. Although hybrid-data and physics-informed approaches have been proposed [12, 13, 14], they often remain computationally expensive and sometimes perform worse than purely data-driven methods [15].

Given the above, this research project aims to address these critical limitations through the following objectives:

1. Enhancing interpretability by approaching FWI from a gradient-flow perspective.
2. Improving the predictive accuracy beyond existing baseline models.
3. Improving generalization to unseen data by reducing dependence on training data. However, the first two objectives will be prioritized.

2 FWI as a gradient flow problem

The inverse problem inherent to FWI can be formulated naturally as a gradient flow problem. Consider the velocity model $u(x)$ as a function evolving continuously in a pseudo-time variable t, denoted as $u(x,t)$. Then equation (4) can be rewritten in continuous time form as:

$$\frac{\partial u}{\partial t} = -\nabla_u \mathcal{J}(p(u)), \tag{5}$$

where the functional $\mathcal{J}(p(u))$ can now be seen as the energy functional. This formulation naturally ensures that the energy $\mathcal{J}$ remains positive and decreases monotonically over time, since

$$\frac{d\mathcal{J}}{dt} = \nabla_u \mathcal{J} \frac{\partial u}{\partial t} = \nabla_u \mathcal{J} \cdot (-\nabla_u \mathcal{J}) = -\|\nabla_u \mathcal{J}\|^2 \leq 0.$$

By adopting this gradient-flow framework, it becomes possible to address the first challenge in 1.1.4 using well-established numerical techniques that guarantee energy dissipation, such as the Scalar Auxiliary Variable (SAV) approach [16]. The idea in this context is to utilize the Evolutional Deep Neural Networks (EDNNs) [17], which explicitly preserves the causal dependencies inherent in time-dependent PDEs, that cannot be guaranteed using conventional neural network methods.

The core motivation behind EDNNs is to determine whether an operator (the PDE solution) for a time-dependent PDE can be accurately learned from only initial input functions (initial conditions). This approach involves parameterizing the initial conditions with a neural network whose parameters are functions of time that evolve continuously. In contrast to conventional neural network approaches, EDNNs enable solution predictions at arbitrary time points and significantly reduce the amount of data required for training.

Extending this approach, [18] applied the EDNN framework, using a DeepONet architecture, to gradient-flow problems while incorporating the SAV technique to ensure consistent energy dissipation over time. This enhancement significantly improves the interpretability of neural operator models. However, a notable drawback of this approach is scalability. The evolution of parameters in EDNN requires solving large linear systems involving all network parameters at every iteration—a computationally prohibitive task, especially for networks with millions of parameters. To mitigate this computational bottleneck, the multi-evolutional deep neural network [19] has been proposed recently.

Utilizing this approach for FWI from a gradient-flow perspective involves first training a neural network (baseline model) to map wavefields directly to subsurface velocities. The parameters of this trained network are then evolved, improving predictions while ensuring continuous energy dissipation. In this way, the output of a well-trained baseline model serves as an initial condition for the evolutionary neural network, directly addressing the second research objective of surpassing existing baseline models in predictive accuracy.

However, an additional critical challenge arises when computing the gradient $\nabla_u \mathcal{J}(p(u))$ needed for the EDNN approach. Notably, this gradient depends explicitly on the velocity u, while the functional $\mathcal{J}$ itself does not directly depend on u. This presents several fundamental questions:

- Can we establish a relationship between $\mathcal{J}(p(u))$ and another functional $E(u)$ that directly depends on u, ensuring a monotonic decrease in $E(u)$ whenever $\mathcal{J}(p(u)$ decreases?

- If such a direct relationship is not feasible, can we circumvent explicit differentiation with respect to u? One viable strategy is the Adjoint state method, a variational approach. However, this approach remains computationally expensive, as it requires solving the forward PDE problem repeatedly at each iteration.

An immediate practical solution to this challenge is employing a machine-learning surrogate for the forward PDE model (1), enabling rapid gradient extraction via automatic differentiation. Although machine-learning surrogates greatly reduce computational cost during inference, they are currently less accurate and less interpretable than traditional numerical PDE solvers, which can induce the cycle-skipping problem if seismic predictions are not accurate.

3 Conclusion

This project aims to address significant challenges inherent in existing machine learning approaches for FWI, specifically their lack of interpretability, limited accuracy, and poor generalization to unseen data. By formulating FWI as a gradient-flow problem, this project aims to enhance interpretability through the explicit consideration of energy dissipation, using innovative techniques such as the SAV method and EDNN. Additionally, by evolving the parameters of initially trained baseline models, we aim to significantly improve prediction accuracy beyond current methodologies. The outcomes of this project are expected to advance the application of SciML methods in seismic inversion, paving the way for more robust and interpretable neural-network-driven inversion strategies.

References

[1] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[2] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[3] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via deeponet based on the universal approximation theorem of operators," *Nature machine intelligence*, vol. 3, no. 3, pp. 218–229, 2021.

[4] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Neural operator: Learning maps between function spaces with applications to pdes," *Journal of Machine Learning Research*, vol. 24, no. 89, pp. 1–97, 2023.

[5] Y. Wu and Y. Lin, "Inversionnet: An efficient and accurate data-driven full waveform inversion," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 419–433, 2019.

[6] Z. Zhang, Y. Wu, Z. Zhou, and Y. Lin, "Velocitygan: Subsurface velocity image estimation using conditional adversarial networks," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 705–714, IEEE, 2019.

[7] M. Zhu, S. Feng, Y. Lin, and L. Lu, "Fourier-deeponet: Fourier-enhanced deep operator networks for full waveform inversion with improved accuracy, generalizability, and robustness," *Computer Methods in Applied Mechanics and Engineering*, vol. 416, p. 116300, 2023.

[8] Z. Guo, L. Chai, S. Huang, and Y. Li, "Inversion-deeponet: A novel deeponet-based network with encoder-decoder for full waveform inversion," *arXiv preprint arXiv:2408.08005*, 2024.

[9] E. Haghighat, U. bin Waheed, and G. Karniadakis, "En-deeponet: An enrichment approach for enhancing the expressivity of neural operators with applications to seismology," *Computer Methods in Applied Mechanics and Engineering*, vol. 420, p. 116681, 2024.

[10] B. Li, H. Wang, S. Feng, X. Yang, and Y. Lin, "Solving seismic wave equations on variable velocity models with fourier neural operator," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.

[11] K. Nath, K. Shukla, V. C. Tsai, U. b. Waheed, C. Huber, O. Alpak, C.-S. Chen, L. Lu, and A. St-Cyr, "Leveraging deep operator networks (deeponet) for acoustic full waveform inversion (fwi)," *arXiv preprint arXiv:2504.10720*, 2025.

[12] M. Rasht-Behesht, C. Huber, K. Shukla, and G. E. Karniadakis, "Physics-informed neural networks (pinns) for wave propagation and full waveform inversions," *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 5, p. e2021JB023120, 2022.

[13] P. Jin, X. Zhang, Y. Chen, S. X. Huang, Z. Liu, and Y. Lin, "Unsupervised learning of full-waveform inversion: Connecting cnn and partial differential equation in a loop," *arXiv preprint arXiv:2110.07584*, 2021.

[14] A. Mardan and G. Fabien-Ouellet, "Physics-informed attention-based neural network for full-waveform inversion," in *85th EAGE Annual Conference & Exhibition (including the Workshop Programme)*, vol. 2024, pp. 1–5, European Association of Geoscientists & Engineers, 2024.

[15] C. Deng, S. Feng, H. Wang, X. Zhang, P. Jin, Y. Feng, Q. Zeng, Y. Chen, and Y. Lin, "Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion," *Advances in Neural Information Processing Systems*, vol. 35, pp. 6007–6020, 2022.

[16] J. Shen, J. Xu, and J. Yang, "The scalar auxiliary variable (sav) approach for gradient flows," *Journal of Computational Physics*, vol. 353, pp. 407–416, 2018.

[17] Y. Du and T. A. Zaki, "Evolutional deep neural network," *Physical Review E*, vol. 104, no. 4, p. 045303, 2021.

[18] J. Zhang, S. Zhang, J. Shen, and G. Lin, "Energy-dissipative evolutionary deep operator neural networks," *Journal of Computational Physics*, vol. 498, p. 112638, 2024.

[19] H. Kim and T. A. Zaki, "Multi evolutional deep neural networks (multi-ednn)," *Journal of Computational Physics*, vol. 531, p. 113910, 2025.