Solving geophysical inverse problems with scientific machine learning

Ziyi Yin
CSE PhD dissertation defense
June 25, 2024

Committee members
Dr. Felix J. Herrmann, advisor, School of CSE, ECE, EAS
Dr. Nisha Chandramoorthy, School of CSE
Dr. Peng Chen, School of CSE
Dr. J. Carlos Santamarina, School of CEE
Dr. Lars Ruthotto, Emory University
Dr. Olav Møyner, SINTEF Digital
Inverse problems

unknown parameters → inverse problem ← indirect measurement
forward operator
Geophysical inverse problems

- unknown Earth parameters
- indirect measurements
- inverse problem
- forward operator
- high-dimensional
High-dimensional parameter estimation

Geophysical exploration and monitoring

- over large subsurface areas
- require high-resolution Earth imaging

\[ nx \times ny \times nz \sim O(10^3 \times 10^3 \times 10^3) \] in realistic settings
Geophysical inverse problems

- Unknown Earth parameters
- Indirect measurements, corrupted by noise
- Forward operator
- Inverse problem

- High-dimensional
Noisy geophysical observations

Weak seismic signals are often corrupted by strong observational noise

Often lead to imaging artifacts
Geophysical inverse problems

- **unknown Earth parameters**
- **indirect measurements**
- **forward operator**
- **inverse problem**

- high-dimensional
- corrupted by noise

- computationally expensive
- non-trivial null-space
- non-convex objective
- non-differentiable legacy simulators
Forward modeling operators
numerical simulators

Computationally expensive

- physics-based simulation
- require solving PDEs

Legacy solvers

- lack interoperability
- difficult to derive sensitivities w.r.t. model parameters

Mathematically challenging

- non-convex objective
- non-trivial null-space
Objectives of my dissertation

Develop scientific machine learning (SciML) methods at scale

- scalable, interoperable, differentiable programming frameworks
- achieve more accurate solutions
- accelerate the inversion process
- provide reliable & affordable inversion
- computationally feasible uncertainty quantification (UQ)
Chapter 2

Learned multiphysics inversion with differentiable programming and machine learning
Motivation
multiphysics inversion

Legacy software

▶ performant, optimized by domain experts — decades of efforts
▶ lack portability & interoperability
▶ difficult to maintain or add new features
▶ (some) lack differentiability & sensitivity calculation

Time-lapse seismic monitoring of geological carbon storage (GCS)

▶ involves coupling of multiphysics modeling & inversion
▶ requires scalable, interoperable & differentiable software stack

hinder R & D
Multiphysics modeling
GCS monitoring

permeability $K$

CO$_2$ saturation $c = \{c_k\}_{k=1}^{n_k}$

seismic velocity $v = \{v_k\}_{k=1}^{n_k}$

time-lapse seismic data $d = \{d_k\}_{k=1}^{n_k}$

fluid-flow physics

rock physics

wave physics

permeability

$K$

CO$_2$ saturation

$v = \{v_k\}_{k=1}^{n_k}$

time-lapse seismic data

$d = \{d_k\}_{k=1}^{n_k}$

Multiphysics modeling
GCS monitoring
Contributions
Chapter 2

Differentiable programming framework via math-inspired software abstractions

- *customized* automatic differentiation (AD) via integration with ChainRules.jl
- coupling of *disjoint* software libraries is feasible and scalable
- easily support *deep learning integration* (e.g., surrogate-assisted inversion)

Case study

- permeability inversion during GCS monitoring
End-to-end inversion framework
multiphysics coupling

permeability
\( K \)

CO\(_2\) saturation
\( c = \{c_k\}_k \)

seismic velocity
\( v = \{v_k\}_k \)

time-lapse
seismic data
\( d = \{d_k\}_k \)

\[ \mathcal{F} \circ \mathcal{R} \circ \mathcal{S} \]
coupled physics

minimize
\( K \)

\[ \left\| \mathcal{F} \circ \mathcal{R} \circ \mathcal{S}(K) - d \right\|^2 \]

End-to-end inversion framework
physics-based

permeability $K$  
$CO_2$ saturation $c = \{c_k\}_{k=1}^{n_k}$  
seismic velocity $v = \{v_k\}_{k=1}^{n_k}$  
time-lapse seismic data $d = \{d_k\}_{k=1}^{n_k}$

Devito / JUDI.jl

JutulDarcy.jl

coupled physics

customized/hand-written

Julia native AD


Møyner, Olav, Grant Bruer, and Ziyi Yin. "Sintefmath/JutulDarcy. jl: V0. 2.3 (version v0. 2.3). Zenodo." (2023).

Yin, Ziyi, Grant Bruer, and Mathias Louboutin. "Slimgroup/JutulDarcyRules. jl: V0. 2.5 (version v0. 2.5). Zenodo." (2023).
End-to-end inversion framework
surrogate-assisted

permeability $K$

$\text{CO}_2$ saturation $c = \{c_k\}_{k=1}^n$

seismic velocity $v = \{v_k\}_{k=1}^n$

data $d = \{d_k\}_{k=1}^n$

End-to-end inversion framework
surrogate-assisted

permeability $K$

$\text{CO}_2$ saturation $c = \{c_k\}_{k=1}^n$

seismic velocity $v = \{v_k\}_{k=1}^n$

data $d = \{d_k\}_{k=1}^n$

customized/hand-written

Julia native AD

$\mathcal{S}_{\theta}^*$ trained Fourier neural operators (FNOs)

$\mathcal{F} \circ \mathcal{R} \circ \mathcal{S}_{\theta}^*$

coupled physics

Case study on the Compass model

**ground truth permeability**

**initial permeability**

**physics-based inversion**

**surrogate-assisted inversion**

---

Chapter 3

Time-lapse full-waveform permeability inversion: a feasibility study
Contributions
Chapter 3

Examine the sensitivities of the permeability inversion framework w.r.t.

- initial model parameters
- modeling errors
- crosstalk during multiparameter inversion

Inversion leads to downstream tasks

- forecast CO$_2$ plume in the future w/o any observation
Chapter 4

Solving multiphysics-based inverse problems with learned surrogates and constraints

Problem formulation

Solve inverse problem: \( \mathbf{d} = \mathcal{H} \circ \mathcal{S}(\mathbf{K}) + \mathbf{\epsilon} \)

- \( \mathbf{d} \) observed data with noise \( \mathbf{\epsilon} \)
- \( \mathbf{K} \) unknown parameter of interest
- \( \mathcal{S} \) modeling operator
- \( \mathcal{H} \) measurement operator
Motivation
surrogate-assisted inversion

minimize \[ \| \mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*}(\mathbf{K}) \|_2^2 \]

Replace numerical simulator \( \mathcal{S} \) by trained FNO \( \mathcal{S}_{\theta^*} \)

- orders of magnitude faster
- auto-differentiable
- intermediate \( \mathbf{K} \) might go out-of-distribution (OOD)
- FNO prediction is less accurate \( \mathcal{S}(\mathbf{K}) \neq \mathcal{S}_{\theta^*}(\mathbf{K}) \)

Goal: “flatten the curve”

FNO error keeps increasing
Propose a learned inversion algorithm

- reap computational benefit of FNO surrogates - fast
- **constrain the FNO input to be always in-distribution** - **accurate**
- still bring down the data misfit via iterative optimization
Normalizing flows (NFs) transport maps

Learn distribution by mapping samples to Gaussian distribution

Mapping by design is **differentiable** and **invertible**


Normalizing flow (for cat)

model space

\[ x \sim p_X(x) \]

latent space

\[ z \sim p_Z(z) \]

training:

\[ G_w^{-1}(x) \]

sampling:

\[ G_w(z) \]
Normalizing flow (for Earth)

training:

\[ \mathbf{K} \sim p_{K}(K) \]  
\[ \mathbf{z} \sim p_{Z}(z) \]

sampling:

\[ \mathbf{z} \sim p_{Z}(z) \]  
\[ \mathbf{K} \sim p_{K}(K) \]
Prior distribution of the Earth shared by FNO & NF training

\[ K^{(1)} \]
\[ K^{(2)} \]
\[ K^{(3)} \]
Invertibility of NF enables probabilistic density evaluation

\[ \mathcal{G}_{w^*}^{-1} \]

model space

\[ \mathcal{G}_{w^*} \]

latent space

in-distribution

out-of-distribution

non-Gaussian for OOD sample
Shrinkage in the latent space of NFs

\[ \ell_2 \text{ norm ball shrinkage} \]

sequence \( \tilde{K} = \mathcal{G}_{w^*}(\alpha z) \) where \( z = \mathcal{G}_{w^*}^{-1}(K) \) and \( 0 \leq \alpha \leq 1 \)

“shrink the latent code and observe the change in the model space”

\[ \alpha = 1 \quad \alpha = 0.4 \quad \alpha = 0.2 \quad \alpha = 0.1 \quad \alpha = 0 \]

in-distribution

out-of-distribution

latent space shrinkage transitions from out-of-distribution to in-distribution
FNO errors during the shrinkage

transitioning from out-of-distribution to in-distribution reduces FNO error
Learned inversion algorithm
with learned surrogates (FNOs) and constraints (NFs)

\[
\begin{align*}
\text{minimize} & \quad \|d - \mathcal{H} \circ S_{\theta^*} \circ G_{w^*}(z)\|^2_2 \\
\text{subject to} & \quad \|z\|_2 \leq \tau
\end{align*}
\]

Trained FNO \( S_{\theta^*} \) replaces numerical simulator \( S \)

Reparameterize the unknown by trained NF \( G_{w^*}(z) \)

\( \tau \) controls size of the iteratively relaxed constraint set

- small \( \tau \) at the beginning ensures to be in-distribution
- gradually increasing \( \tau \) brings down the objective
Permeability inversion results
unconstrained inversion with FNO surrogates

\[
\minimize_{\mathbf{K}} \| \mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*}(\mathbf{K}) \|_2^2
\]

visible artifacts in the recovery
Permeability inversion results
constrained inversion with FNO surrogates

\[
\begin{align*}
\text{minimize} \quad & \|d - \mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{w^*}(z)\|_2^2 \\
\text{subject to} \quad & \|z\|_2 \leq \tau
\end{align*}
\]

NF constraint greatly improves inversion
FNO error along iterations
constrained vs unconstrained inversion

seismic observations

FNO error [%]
0 8 10 12 14 16
0 10 20 30 40 50
iterations

seismic + well observations

FNO error [%]
0 8 10 12 14
0 10 20 30 40 50
iterations

FNO error remains relatively flatline during constrained inversion
Conclusions & Contributions
Chapter 4

After training FNO & NF on the same samples

- FNO error can be controlled by latent space shrinkage of NF

Propose learned inversion algorithm with FNO & NF

- NF reparameterization forms an efficient continuation scheme / homotopy
- iteratively relaxed constraint
  - safeguard FNO accuracy
  - bring down objective

Proof-of-concept permeability inversion from time-lapse seismic + well data
Chapter 5

Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection
Contributions
Chapter 5

Propose low-cost time-lapse seismic acquisition & imaging

Monitor CO$_2$ dynamics when it *fails to follow* multiphase flow equations

Deploy the joint recovery model (JRM)

  ▶ exploit *shared information* to enhance imaging quality

  ▶ reduce reliance on *replicating* source & receiver positions across surveys

Train deep neural classifiers

  ▶ automatic leakage detection from time-lapse seismic images

  ▶ explainable saliency maps
Chapter 6

WISE: full-Waveform variational Inference via Subsurface Extensions
Geophysical exploration

velocity

\[ x \]

\[ \mathcal{F} \]

wave physics

seismic data

\[ y \]
Full-waveform inversion (FWI)
Inverse problems related to PDE parameter estimation

velocity

\( x \)

\( \mathcal{F} \)

wave physics

seismic data

\( y \)
FWI cont’d

\[ y = \mathcal{F}(x) + \epsilon \]

- **x**: acoustic velocity (unknown parameter of interest)
- **\( \mathcal{F} \)**: nonlinear forward modeling operator
- **y**: observed seismic data
- **\( \epsilon \)**: noise
Bayesian inference
posterior

\[ p( x \mid y ) \]

unknown parameter

observed data
Full-waveform inversion & inference
posterior

$$p(\text{velocity model } x \mid \text{observed data } y)$$
Amortized variational inference (VI)

Learn $q_\theta(x \mid y) \approx p(x \mid y)$ via sample pairs $\{x^{(i)}, y^{(i)}\}_{i=1}^N$

Train conditional normalizing flows (CNFs)

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{2} \| f_\theta(x^{(i)}; y^{(i)}) \|_2^2 - \log | \det J_{f_\theta} | \right)$$

- $p$ unknown target posterior distribution
- $q_\theta$ approximated posterior distribution via CNFs $f_\theta$
- expensive offline training
- cheap online inference
Challenges

VI w/ CNFs

Practical challenges of training CNFs to approximate $p(x \mid y)$

- need to *retrain* for new configurations (e.g., source/receiver positions)
- mapping between image $x$ and data $y$ is very *difficult to learn*
- does not incorporate any *physics* during training & inference

Current literature suggests

- *physics-informed summary statistics*
- partially *decode* the wave physics
Full-waveform inference approximated posterior

\[ p(x | \bar{y}) \]

velocity model \( x \)  
summary statistics \( \bar{y} \)

Summary statistics need to
- preserve all information in data
- decode the complicated wave physics
Motivation
model extension & extended gradients

Orozco et al proved for linear inverse problems

- \( y = Ax + \epsilon \)  where  \( \epsilon \sim N(0, I) \)

- \( p(x \mid y) \equiv p(x \mid \bar{y}) \)  where  \( \bar{y} = A^\top y \)

Linearize FWI problem at velocity  \( x_0 \)

- \( \mathcal{F}(x) \approx \mathcal{F}(x_0) + \nabla \mathcal{F}(x_0)(x - x_0) \)

Consider Gauss-Newton update at a bad linearization point  \( x_0 \)

- standard Jacobian can’t drive residual to 0, “information is lost”

- extended Jacobian can preserve information

\[
\begin{align*}
\text{standard} & \quad \min_{\delta x} \| \delta y - \nabla \mathcal{F}(x_0) \delta x \|_2^2 \\
\text{extended} & \quad \min_{\delta x} \| \delta y - \nabla \mathcal{F}(x_0) \delta x \|_2^2
\end{align*}
\]

Correct:  \( x_0 \) is close to  \( x \)
Wrong:  \( x_0 \) is far from  \( x \)

Standard gradient

\[ \mathbf{g}[\mathbf{x}] = \nabla \mathcal{F}(\mathbf{x}_0)^T \delta \mathbf{y} = \sum_{i=1}^{n_t} \sum_{t=1}^{n_s} \mathbf{u}_i[\mathbf{x}, t] \odot \mathbf{v}_i[\mathbf{x}, t] \]

- \( \mathbf{u}_i[\mathbf{x}, t] \): second-time derivative solution of wave equation: \( \mathbf{A}(\mathbf{x}_0) \mathbf{u}_i = \mathbf{q}_i \)

- \( \mathbf{v}_i[\mathbf{x}, t] \): solution of adjoint wave equation: \( \mathbf{A}(\mathbf{x}_0)^T \mathbf{v}_i = \mathbf{P}_r^T \delta \mathbf{y}_i \)

Extended gradient (with an extra subsurface-offset dimension)

\[ \mathbf{g}[\mathbf{x}, \mathbf{h}] = \overline{\nabla \mathcal{F}(\mathbf{x}_0)}^T \delta \mathbf{y} = \sum_{i=1}^{n_t} \sum_{t=1}^{n_s} \mathbf{u}_i[\mathbf{x} + \mathbf{h}, t] \odot \mathbf{v}_i[\mathbf{x} - \mathbf{h}, t] \]

- note: \( \mathbf{g}[\mathbf{x}] = \mathbf{g}[\mathbf{x}, \mathbf{0}] \)

- near isometry & acts as an embedding
Extended gradient

good $x_0$
ML4Seismic

Extended gradient

poor $x_0$

preserve information at non-zero offsets
Full-waveform inference posterior

\[ p(\text{velocity model } x \mid \text{observed data } y) \]

Mapping between velocity model and data is very difficult to learn
Full-waveform inference
summary statistics = extended gradient

\[ p(\text{velocity model } x, \text{extended gradient } \bar{y}) \]

decode wave physics & preserve information
Unseen ground truth velocity
Conditional mean estimate
summary statistics = standard gradient
Conditional mean estimate
summary statistics = extended gradient
Uncertainty in imaged reflectivities entails important information to make business decisions.
Amplitude variations point-wise standard deviation
Positioning variations
maximum vertical shift via cross correlation
Contributions
Chapter 6

Propose physics-informed summary statistics for *nonlinear FWI* problem

- based on *model extension* and geophysical knowledge
- reduce reliance on *accurate* initial model
- preserve information
- enhance CNF training

Perform *forward UQ* for downstream imaging tasks
Chapter 7

WISER: multimodal variational inference for full-waveform inversion without dimensionality reduction

WISER = WISE + Refinements
based on wave physics

Challenges

▶ amortization gap
- network works well for a family of observations
- but does not provide very accurate prediction for a single observation
▶ out of distribution at inference

Solution

▶ fine-tune network via a few physics-based iterations
Physics-based latent space correction constrained formulation

\[
\begin{align*}
\min_{\phi} \ & \ \mathbb{K}L \left( p \left( h_\phi(z) \right) \mid p_{\text{post}}(z \mid \overline{y}_{\text{obs}}) \right) \\
& = \mathbb{E}_{z \sim \mathcal{N}(0, I)} \left[ \frac{1}{2\sigma^2} \| \mathcal{F} \circ f_{\theta^*}^{-1} \left( h_\phi(z); \overline{y}_{\text{obs}} \right) - \overline{y}_{\text{obs}} \|_2^2 + \frac{1}{2} \| h_\phi(z) \|_2^2 - \log \left| \det J_{h_\phi} \right| \right] \\
\end{align*}
\]

\[f_\theta\] trained \textit{amortized} CNF from WISE, \(h_\phi\) refined \textit{non-amortized} NF

Challenge: physics \(\mathcal{F}\) (expensive) and networks \(f_{\theta^*}\), \(h_\phi\) are always coupled

\textbf{Solution: decouple them via \textit{weak} formulation}
Proposed WISER objective

weak formulation

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{M} \sum_{i=1}^{M} \left[ \frac{1}{2\sigma^2} \| \mathcal{F}(x_i) - y_{\text{obs}} \|_2^2 + \frac{1}{2\gamma^2} \| x_i - f_{\theta^*}^{-1}(h_{\phi}(z_i); y_{\text{obs}}) \|_2^2 + \frac{1}{2} \| h_{\phi}(z_i) \|_2^2 \right. \\
& \quad \left. - \log \left| \det J_{h_{\phi}} \right| \right] \\
\end{align*}
\]

\begin{align*}
\text{likelihood} & \quad \text{weak prior} & \quad \text{prior}
\end{align*}

When $\gamma \to 0$, weak formulation $\to$ constrained formulation

**Outer loop:** update $x_{1:M}$ using expensive physics $\mathcal{F}$ — a few times

**Inner loop:** update $\phi$ using only networks — many times
Distribution shift at inference

in distribution

out of distribution

element-wise perturbation
Predicted velocity models - WISE
Predicted velocity models - WISER
Histograms

Depth = 0.5 km

Depth = 2.8 km

WISE is close to original
WISER is close to perturbed
Imaged reflectivities before correction - WISE

Layers are disconnected
Imaged reflectivities after correction - WISER

layers are connected and aligned with the velocity model
Propose physics-based refinement approach to improve WISE

- *frugal* usage of wave modeling and gradient
- robust w.r.t. *OOD scenarios* at inference

WISER leads to a novel *semi-amortized VI* paradigm

- computationally *affordable & scalable*
- physics-based & *reliable*
- not local, but *global* optimization & UQ
- w/o dimensionality reduction
Summary of contributions

Design *interoperable* and *differentiable* programming framework to support learned multiphysics inversion at scale.

Explore deep neural networks as surrogate models to learn:
- forward map
  - safeguard the accuracy of surrogate simulators during inversion via *learned constraints*
- (nonunique) inverse map
  - *physics-informed* & *information-preserving* summary statistics based on extension of wave physics
  - mitigate the amortization gap via *affordable physics-based refinements*

Employ the proposed SciML algorithm to solve inverse problems that are:
- *high-dimensional*
- with *computationally expensive* forward operators

Including:
- full-waveform inversion
- geological carbon storage monitoring
Future directions

Surrogate-assisted inversion with learned constraints
  ▶ examine different parameterizations
  ▶ derivative-informed surrogate-assisted inversion

Semi-amortized VI w/ WISE & WISER
  ▶ theoretically explore the family of model-extension-based summary statistics
  ▶ choice of initial model / fiducial point
    - experimental configuration in Bayesian optimal experimental design
    - nuisance parameter in simulation-based inference
  ▶ more challenging distribution of model parameters (salt bodies) & OOD scenarios
**Journal papers**


Many thanks to

Felix J. Herrmann
PhD proposal & dissertation defense committee members
Colleagues at SLIM
Olav Møyner at SINTEF
Rishi Khan at Extreme Scale Solution
Internship colleagues at Chevron
NSF, DOE, ML4SEISMIC center, Georgia Research Alliance
Family and friends