Learning by example: fast reliability-aware seismic imaging with normalizing flows

Ali Siahkoohi 1  Gabrio Rizzuti 2  Philipp A. Witte 3  Mathias Louboutin 1  Felix J. Herrmann 1

1 Georgia Institute of Technology
2 Utrecht University
3 Microsoft Research for Industry
Motivation

Seismic imaging is challenged by

- noisy data and linearization errors
- bandwidth and aperture limitations
- presence of shadow zones
- expensive-to-evaluate forward operator

Leads to solution non-uniqueness

- different images may fit the data equally well

Failure to assess uncertainty may have implications on identification of risk

- risk associated with tasks: horizon tracking, semantic segmentation, etc
- challenged by
  - high-dimensionality of seismic images
  - expensive-to-evaluate migration/demigration operator
Problem setup

Find \( x \in \mathcal{X} \) such that

\[
y = F(x) + n
\]

- expensive (nonlinear) forward operator \( F \)
- unknown quantity \( x \)
- observed data \( y \in \mathcal{Y} \)
- (non-Gaussian) noise \( n \)

Bayesian inference

\[
p_{\text{post}}(x \mid y) \propto p_{\text{like}}(y \mid x) \, p_{\text{prior}}(x)
\]

---

Markov chain Monte Carlo (MCMC) sampling

\[ x_{k+1} = x_k + \frac{\alpha_k}{2} \nabla_x \left[ \log p_{\text{like}}(y \mid x_k) + \log p_{\text{prior}}(x_k) \right] + \xi_k, \quad \xi_k \sim \mathcal{N}(0, \alpha_k I) \]

- asymptotically exact samples
- high-dimensional integration/sampling
- costs of forward operator
- requires choosing a prior distribution


Variational Inference

Approximate target density $p_{\text{post}}(x \mid y)$ via parametric density $p_{\theta}(x \mid y)$

$$\forall x, y : \quad p_{\theta}(x \mid y) \approx p_{\text{post}}(x \mid y)$$

- admits stochastic optimization
- cheap sampling after optimization
- emerging evidence that scales better than MCMC


Normalizing flows (NFs)

Invertible neural network $G_\theta : \mathcal{X} \to \mathcal{Z}$, $\mathcal{X}, \mathcal{Z} \in \mathbb{R}^d$

- flexible function for variational inference
- Gaussian latent space $p_z(z) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}\|z\|^2_2}$
- closed-from and exact inverse (up to numerical precision)
- memory-efficient training due to invertibility
- closed-form expression for $p_\theta(x)$

$$p_\theta(x) = p_z(G_\theta(x)) \left|\det \nabla_x G_\theta(x)\right| \approx p_X(x)$$

Data-driven training of NFs

\[ \begin{align*}
\arg\min_{\theta} \mathbb{D}_{KL}(p_X \mid \mid p_\theta) &= \arg\min_{\theta} \mathbb{E}_{x \sim p_X(x)} \left[ -\log p_\theta(x) + \log p_X(x) \right] \\
&\approx \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{1}{2} \| G_\theta(x_i) \|_2^2 - \log \left| \det \nabla_x G_\theta(x_i) \right| \right]
\end{align*} \]

- **Kullback-Leibner (KL) divergence** \( \mathbb{D}_{KL} \)
- **training samples** \( \{x^{(i)}\}_{i=1}^{n} \sim p_X(x) \)
- **\( \det \nabla_x G_\theta(x) \)** comes for free with **affine coupling layers**


Conditional NFs

\[
\min_\theta \mathbb{E}_{y, x \sim p(y, x)} \left[ \frac{1}{2} \left\| G_\theta(y, x) \right\|^2 - \log \left| \det \nabla_{y, x} G_\theta(y, x) \right| \right]
\]

\[G_\theta(y, x) = \begin{bmatrix} G_{\theta_y}(y) \\ G_{\theta_x}(y, x) \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_y \\ \theta_x \end{bmatrix}\]

- conditional NF, \(G_\theta : \mathcal{Y} \times \mathcal{X} \rightarrow \mathcal{Z}_y \times \mathcal{Z}_x\)
- expectation estimated via training pairs, \(\{y_i, x_i\}_{i=1}^n \sim p_{Y, X}(y, x)\)
Posterior inference with conditional NFs

- cheap samples from posterior (no PDE solves besides one RTM to get $y$)

\[
G_{\theta_x}^{-1}(G_{\theta_y}(y), z_x) \sim p_{\text{post}}(x \mid y), \quad z_x \sim N(0, I)
\]

- cheap posterior density estimation

\[
p_{\theta}(x \mid y) = p_z(G_{\theta_x}(y, x)) \left| \det \nabla_x G_{\theta_x}(y, x) \right| \approx p_{\text{post}}(x \mid y)
\]

Seismic imaging example setup

\[ d = J[m_0, q] \delta m + \eta + n, \quad n \sim N(0, \sigma^2 I) \]

- linearized observed data \( d \)
- linearized Born demigration operator \( J \)
- source signature \( q \)
- smooth background model \( m_0 \)
- unknown perturbation model \( \delta m \)
- linearization error and noise \( \eta, n \)

training pairs

\[ \{y_i, x_i\}_{i=1}^n \sim p(y, x) \]

where

\[ (y_i, x_i) = (J^\top (J \delta m_i + n_i), \delta m_i) \]
Low- and high-fidelity images

For training, we select $\delta m$ as

- $3075 \text{ m} \times 5120 \text{ m}$ sections from the Parihaka dataset
- artificial 125 m water layer to limit the near source imaging artifacts

Other possible choices

- pairs of RTM and LS-RTM images
- migrated images with migration swings and postprocessed clean migrated images
- migrated images with and without errors in background model, ...

---

High-fidelity image, \( \mathbf{x} \)
Low-fidelity reverse-time migrated image, $y$
Conditional mean, $\delta m_{CM}$

\[ \mathbb{E}[x | y] = \int x p_{post}(x | y) \, dx \]
\[ \mathbb{E}\left[ (\mathbf{x} - \mathbb{E}[\mathbf{x} | \mathbf{y}])^2 | \mathbf{y} \right] = \int (\mathbf{x} - \mathbb{E}[\mathbf{x} | \mathbf{y}])^2 p_{\text{post}}(\mathbf{x} | \mathbf{y}) \, d\mathbf{x} \]
Contributions

Learning by example: maximally leveraging existing RTM and LS-RTM images

Sampling from posterior virtually for free (no PDE solves), providing
- cheap access to the high-fidelity image, comparable to LS-RTM
- a rudimentary assessment of its uncertainty
- orders of magnitude faster Bayesian inference compared to traditional methods

Applicable to previously unseen seismic data from neighboring surveys

Can be later tied to physics for more accurate inference...
Multifidelity preconditioned formulations

- insert the conditional NF in physics-based formulations and use transfer learning

\[ T_{\theta_x}(z) := G_{\theta_x}^{-1}(G_{\theta_y}(y), z) \]

- preconditioned maximum a posteriori (MAP) estimation

\[ \min_z \frac{1}{2\sigma^2} \| F(T_{\theta_x}(z)) - y \|_2^2 + \frac{1}{2} \| z \|_2^2 \]

- preconditioned physics-based posterior learning

\[ \min_{\theta_x} \mathbb{E}_{z \sim \mathcal{N}(0, I)} \left[ \frac{1}{2\sigma^2} \| F(T_{\theta_x}(z)) - y \|_2^2 - \log p_{\text{prior}}(T_{\theta_x}(z)) - \log \left| \det \nabla_z T_{\theta_x}(z) \right| \right] \]

Conclusions

Obtaining UQ information is impractical when

- the forward operators are expensive to evaluate
- the problem is high dimensional

Bayesian inference with normalizing flows

- fast and cheap access to the high-fidelity conditional mean estimate and a first assessment of uncertainty
- precondition a physics-based inference with a pretrained conditional normalizing flow

Future work

- characterize uncertainties due to errors in modeling and background model
Code

https://github.com/slimgroup/InvertibleNetworks.jl

https://github.com/slimgroup/Software.SEG2021

Acknowledgment

This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center