Wave-equation-based inversion with amortized variational Bayesian inference

Ali Siahkoohi ¹  Gabrio Rizzuti ²  Rafael Orozco ¹  Felix J. Herrmann ¹

¹Georgia Institute of Technology
²Utrecht University
Importance of uncertainty quantification

Seismic imaging is challenged by

- noisy data and linearization errors
- bandwidth and aperture limitations
- presence of shadow zones

Failure to assess uncertainty may have implications on downstream tasks

- uncertainty associated with tasks: horizon tracking, semantic segmentation, etc
- challenged by
  - high-dimensionality of seismic images
  - expensive-to-evaluate migration/demigration operator
Challenges of solving Bayesian inverse problems

Choosing a prior distribution

- encode prior knowledge
- avoid unwanted bias due to overly simplifying priors

Computational cost

- costly forward operator
- high dimensional sampling/integration
Deep generative networks for solving inverse problems

- learned prior and posterior distributions
- fast conditional sampling
- tractable density estimation
- rely on access to high-quality training data
- negative bias induced by distribution shifts during inference


Proposed approach

learning prior and amortized posterior distributions with conditional normalizing flows

data-specific (non-amortized), low-cost, physics-based latent distribution correction

► cheap and unlimited posterior samples
► directly informed by data and physics
► minimizes the negative bias of distribution shifts during inference
► feasible in domains with limited access to training data
Amortized posterior sampling w/ normalizing flows

\[ \mathcal{N}(\mathbf{z} \mid 0, \mathbf{I}) \]

\[ \mathbf{y}_{\text{obs}} \sim p_{\text{data}}(\mathbf{y}) \]

\[ \mathbf{x} \sim p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}}) \]


Simulation-based inference

\[ x, y \sim p(x, y) \]

\[ y \sim p_{\text{like}}(y \mid x) \]

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

\[ x \sim p_{\text{prior}}(x) \]

\[ f_\phi(x; y) \]

\[ \min_\phi \mathbb{E}_{x, y \sim p(x, y)} \left[ \frac{1}{2} \left\| f_\phi(x; y) \right\|_2^2 - \log \left| \det \nabla_x f_\phi(x; y) \right| \right] \]

\[ N(z \mid 0, I) \]
Seismic imaging example

in-distribution amortized posterior sampling
Reverse-time migrated image

SNR $-12.17 \text{ dB}$
Posterior sample (amortized)

SNR 8.54 dB
Posterior sample (amortized)

Depth (km)

Horizontal distance (km)

SNR 8.49 dB
Posterior sample (amortized)

SNR 8.42 dB
Posterior sample (amortized)

SNR 8.59 dB
Conditional mean estimate (amortized)

SNR 11.24 dB
Normalized pointwise standard deviation (amortized)

normalized by the envelope of the conditional mean
Introducing distribution shifts

band-limited noise with $6.25 \times$ larger variance

$4 \times$ less sources

Physics-based latent distribution correction

computational cost: approximately $5 \times$ RTMs
Reverse-time migrated image

SNR $-8.22$ dB
Ground-truth (unknown) image

Previously unseen (test) seismic image
Physics-based latent distribution correction

\[ N(z \mid 0, I) \]

\[ N(z \mid \mu, \text{diag}(s)^2) \]

\[ p_{\text{post}}(x \mid y_{\text{obs}}) \]

\[ \hat{p}_{\text{data}}(y) \]

\[ y_{\text{obs}} \sim \hat{p}_{\text{data}}(y) \]
Physics-based latent distribution correction

\[
\arg \min_{\mu, s} \mathbb{E}_{z \sim N(z|0, I)} \left[ \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\| y_{\text{obs},i} - \mathcal{F} \circ f_{\phi}^{-1} (s \odot z + \mu; y_{\text{obs}}) \right\|_2^2 + \frac{1}{2} \left\| s \odot z + \mu \right\|_2^2 - \log \det \text{diag}(s) \right]
\]
Normalized pointwise standard deviation (corrected)

normalized by the envelope of the conditional mean
Data space QC

inspect data residual before and after latent distribution correction
(left) noise free data (right) noisy data with SNR $-2.78$ dB
predicted data (left) amortized, SNR 11.62 dB (right) corrected, SNR 16.57 dB
data residual of (left) amortized (right) corrected
Multi-experiment inverse problems

Find \( \mathbf{x} \) such that

\[
\mathbf{y}_i = \mathcal{F}_i(\mathbf{x}) + \mathbf{\epsilon}_i, \quad \mathbf{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad i = 1, \ldots, N
\]

- observed data \( \mathbf{y} = \{\mathbf{y}_i\}_{i=1}^N, \; \mathbf{y}_i \in \mathcal{Y} \)
- unknown quantity \( \mathbf{x} \in \mathcal{X} \)
- expensive-to-evaluate forward operator \( \mathcal{F}_i : \mathcal{X} \to \mathcal{Y} \)
- noise and/or modeling error \( \mathbf{\epsilon}_i \)
- noise covariance \( \sigma^2 \mathbf{I} \)
Negative-log posterior

\[- \log p_{\text{post}}(x \mid y) = - \sum_{i=1}^{N} \log p_{\text{like}}(y_i \mid x) - \log p_{\text{prior}}(x) + \log p_{\text{data}}(y)\]

\[= \frac{1}{2\sigma^2} \sum_{i=1}^{N} \|y_i - \mathcal{F}_i(x)\|_2^2 - \log p_{\text{prior}}(x) + \text{const.}\]

likelihood function following Gaussian noise assumption

prior distribution \(p_{\text{prior}}(x)\)

constant term (data density) independent of \(x\)
Amortized variational inference with normalizing flows

\[ \phi^* = \arg \min_{\phi} \mathbb{E}_{y \sim p_{data}(y)} \left[ \mathbb{KL} \left( p_{post}(x \mid y) \parallel p_{\phi}(x \mid y) \right) \right] \]

\[ = \arg \min_{\phi} \mathbb{E}_{x, y \sim p(x, y)} \left[ \frac{1}{2} \left\| f_{\phi}(x; y) \right\|_2^2 - \log \left| \det \nabla_x f_{\phi}(x; y) \right| \right] \]

normalizes the input

entropy regularization
e.g., avoids \( f_{\phi}(x; y) \equiv 0 \)

\( f_{\phi}(\cdot; y) : \mathcal{X} \rightarrow \mathcal{Z} \) an invertible neural net

approximate expectation with available model and data pairs \( \{(x_i, y_i)\}_{i=1}^n \sim p(x, y) \)

negligible computational cost of \( \det \nabla_x f_{\phi}(x; y) \)'s gradient due to \( f_{\phi} \)'s architecture

Amortized variational inference—sampling

For previously unseen data $y_{obs} \sim p_{data}(y)$

$$f_{\phi}^{-1}(z; y_{obs}) \sim p_{post}(x \mid y_{obs}), \quad z \sim N(0, I)$$

learned prior and posterior distributions

fast and cheap amortized posterior sampling given new observation $y_{obs} \sim p_{data}(y)$

no optimization or MCMC sampling for new observations

---


Implications of imperfect pretraining

Inaccurate normalization through the conditional normalizing flow

- $p_\phi(z \mid y_{\text{obs}}) \neq N(z \mid 0, I)$

- $p_\phi(z \mid y_{\text{obs}})$ distribution of $z = f_\phi(x; y_{\text{obs}})$ for $x \sim p_{\text{post}}(x \mid y_{\text{obs}})$, $y_{\text{obs}} \sim p_{\text{data}}(y)$

Inaccurate posterior sampling given $z \sim N(z \mid 0, I)$ as input

- $p_\phi(x \mid y_{\text{obs}}) \neq p_{\text{post}}(x \mid y_{\text{obs}})$

- $p_\phi(x \mid y_{\text{obs}})$ predicted posterior distribution

---

Latent distribution correction

\[ z \sim N(z \mid 0, I) \quad \rightarrow \quad z \sim N(z \mid \mu, \text{diag}(s)^2) \]

For certain classes of distribution shifts

- \( p_\phi(z \mid y) \) not too far from normal distribution

- improving the initial approximation \( p_\phi(z \mid y) \approx N(z \mid 0, I) \)
Physics-based latent distribution correction

\[
\min_{\mu, s} \mathbb{E}_{z \sim \mathcal{N}(0,1)} \left[ \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\| y_{\text{obs},i} - \mathcal{F}_i \circ f^{-1}_\phi (s \odot z + \mu; y_{\text{obs}}) \right\|^2_2 + \frac{1}{2} \left\| s \odot z + \mu \right\|^2_2 - \log \left| \det \text{diag}(s) \right| \right]
\]

initializing with \( \mu = 0 \) and \( \text{diag}(s)^2 = I \)

initialization acts as a warm-start and an implicit regularization

expected to be solved relatively cheaply due to the amortization of \( f_\phi \)

non-amortized, i.e., specific to one set of observations \( y_{\text{obs}} \sim \hat{p}_{\text{data}}(y) \)
Non-amortized latent distribution correction

\[
\arg\min_{\mu, s} \mathbb{E}_{z \sim \mathcal{N}(z|0, I)} \left[ \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\| \mathbf{y}_{\text{obs},i} - \mathcal{F} \circ f^{-1}_\phi (s \odot \mathbf{z} + \mathbf{\mu}; \mathbf{y}_{\text{obs}}) \right\|^2_2 + \frac{1}{2} \left\| s \odot \mathbf{z} + \mathbf{\mu} \right\|^2_2 - \log \left| \det \text{diag}(s) \right| \right]
\]
Conclusions

Uncertainty quantification is rendered impractical when
- the forward operators are expensive to evaluate
- the problem is high dimensional

Amortized variational inference with physics-based latent distribution correction
- can lead to orders of magnitude computational improvements compared to MCMC and traditional variational inference methods
- limits the adverse affects of data distribution shifts
- provides fast (same cost as 5 RTMs) and reliable posterior inference
Code

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

https://github.com/slimgroup/ConditionalNFs4Imaging.jl
Acknowledgment

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