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

SLIM
Georgia Institute of Technology

Motivation

Seismic imaging is challenged by

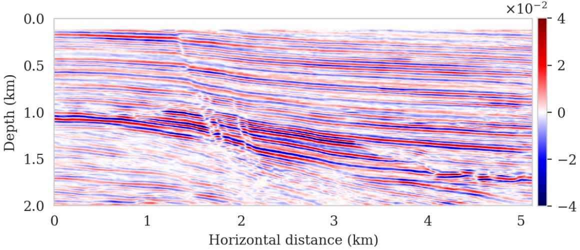
- noisy data and linearization errors
- bandwidth and aperture limitations
- presence of shadow zones
- expensive-to-evaluate forward operator

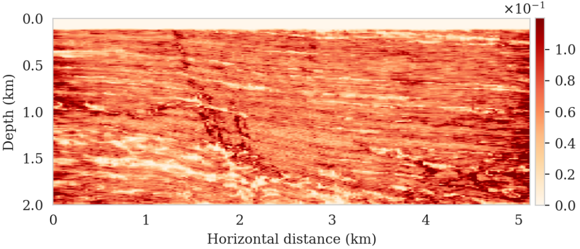
Leads to solution non-uniqueness

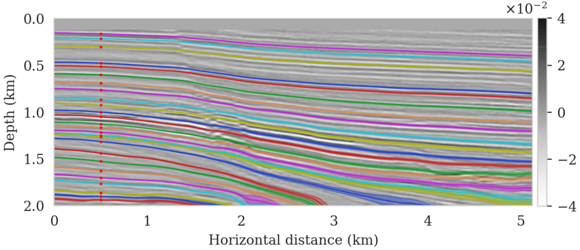
- different images may fit the data equally well

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







Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. “Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach”. In: *SEG Technical Program Expanded Abstracts 2020*. Sept. 2020. DOI: 10.1190/segam2020-3417560.1.

Problem setup

Find $\mathbf{x} \in \mathcal{X}$ such that

$$\mathbf{y} = F(\mathbf{x}) + \mathbf{n}$$

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

Bayesian inference

$$p_{\text{post}}\left(\mathbf{x} \mid \mathbf{y}\right) \propto p_{\text{like}}(\mathbf{y} \mid \mathbf{x})\, p_{\text{prior}}(\mathbf{x})$$

Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005. ISBN: 978-0-89871-572-9. DOI: 10.1137/1.9780898717921.

Markov chain Monte Carlo (MCMC) sampling

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\alpha_k}{2}\nabla_{\mathbf{x}}\Big[\log p_{\text{like}}(\mathbf{y} \mid \mathbf{x}_k) + \log p_{\text{prior}}(\mathbf{x}_k)\Big] + \boldsymbol{\xi}_k, \quad \boldsymbol{\xi}_k \sim \mathrm{N}(\mathbf{0}, \alpha_k \mathbf{I})$$

- asymptotically exact samples

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

Christian Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2004. DOI: 10.1007/978-1-4757-4145-2.

Max Welling and Yee Whye Teh. "Bayesian Learning via Stochastic Gradient Langevin Dynamics". In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. ICML'11. 2011, pp. 681–688. DOI: 10.5555/3104482.3104568.

Variational Inference

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

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

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

- requires samples from $p_{\text{prior}}(\mathbf{x})$ or analytical density $p_{\text{prior}}(\mathbf{x})$

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. "Variational inference: A review for statisticians". In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. "An Introduction to Variational Methods for Graphical Models". In: *Machine Learning* 37.2 (1999), pp. 183–233. DOI: 10.1023/A:1007665907178.

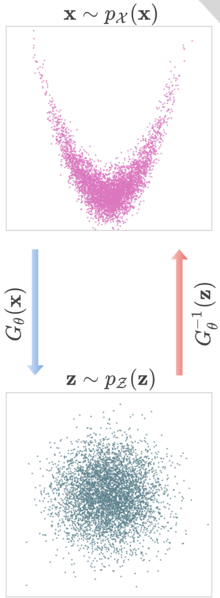
Xin Zhang and Andrew Curtis. "Seismic tomography using variational inference methods". In: *Journal of Geophysical Research: Solid Earth* 125.4 (2020), e2019JB018589.

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

$$p_\theta(\mathbf{x}) = p_z(G_\theta(\mathbf{x})) \left| \det \nabla_x G_\theta(\mathbf{x}) \right| \approx p_{\mathcal{X}}(\mathbf{x})$$



Danilo Rezende and Shakir Mohamed. "Variational Inference with Normalizing Flows". In: vol. 37. Proceedings of Machine Learning Research. PMLR, 2015, pp. 1530–1538. URL: http://proceedings.mlr.press/v37/rezende15.html.

Data-driven training of NFs

$$\underset{\boldsymbol{\theta}}{\arg\min}\ \mathbb{D}_{\mathsf{KL}}\left(p_{\mathcal{X}} \mid\mid p_{\theta}\right) = \underset{\boldsymbol{\theta}}{\arg\min}\ \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{X}}(\mathbf{x})}\Big[-\log p_{\theta}(\mathbf{x}) + \log p_{\mathcal{X}}(\mathbf{x})\Big]$$

$$\approx \underset{\theta}{\arg\min}\ \frac{1}{n}\sum_{i=1}^{n}\Bigg[\underbrace{\frac{1}{2}\left\|G_{\theta}(\mathbf{x}_i)\right\|_2^2}_{\text{Gaussianizes the input}} - \underbrace{\log\left|\det \nabla_x G_{\theta}(\mathbf{x}_i)\right|}_{\substack{\text{regularizes entropy of } p_{\theta}(\mathbf{x}) \\ \text{e.g., avoids } p_{\theta}(\mathbf{x}) = \delta_{\bar{\mathbf{x}}}(\mathbf{x})}}\Bigg]$$

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

Danilo Rezende and Shakir Mohamed. "Variational Inference with Normalizing Flows". In: vol. 37. Proceedings of Machine Learning Research. PMLR, 2015, pp. 1530–1538. URL: http://proceedings.mlr.press/v37/rezende15.html.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using Real NVP". In: *International Conference on Learning Representations, ICLR*. 2016. URL: http://arxiv.org/abs/1605.08803.

Conditional NFs

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

$$G_{\theta}(\mathbf{y},\mathbf{x}) = \begin{bmatrix} G_{\theta_y}(\mathbf{y}) \\ G_{\theta_x}(\mathbf{y},\mathbf{x}) \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_y \\ \boldsymbol{\theta}_x \end{bmatrix}$$

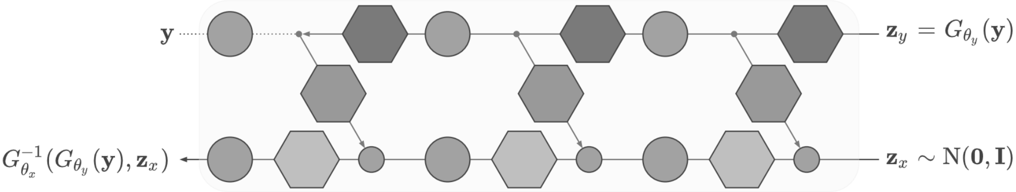
- conditional NF, $G_{\theta} : \mathcal{Y} \times \mathcal{X} \to \mathcal{Z}_y \times \mathcal{Z}_x$
- expectation estimated via training pairs, $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^{n} \sim p_{\mathcal{Y},\mathcal{X}}(\mathbf{y},\mathbf{x})$

Jakob Kruse, Gianluca Detommaso, Robert Scheichl, and Ullrich Köthe. "HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference". In: *Proceedings of AAAI-2021* (2021). URL: https://arxiv.org/pdf/1905.10687.pdf.

Ali Siahkoohi, Gabrio Rizzuti, Mathias Louboutin, Philipp Witte, and Felix J. Herrmann. "Preconditioned training of normalizing flows for variational inference in inverse problems". In: *3rd Symposium on Advances in Approximate Bayesian Inference*. Jan. 2021. URL: https://openreview.net/pdf?id=P9m1sMaNQ8T.

Ali Siahkoohi and Felix J. Herrmann. "Learning by example: fast reliability-aware seismic imaging with normalizing flows". Apr. 2021. URL: https://arxiv.org/pdf/2104.06255.pdf.

Posterior inference with conditional NFs



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

$$G_{\theta_x}^{-1}(G_{\theta_y}(\mathbf{y}), \mathbf{z}_x) \sim p_{\text{post}}(\mathbf{x} \mid \mathbf{y}), \quad \mathbf{z}_x \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$$

- cheap posterior density estimation

$$p_\theta(\mathbf{x} \mid \mathbf{y}) = p_{\mathbf{z}}\big(G_{\theta_x}(\mathbf{y}, \mathbf{x})\big) \left| \det \nabla_x G_{\theta_x}(\mathbf{y}, \mathbf{x}) \right| \approx p_{\text{post}}(\mathbf{x} \mid \mathbf{y})$$

Jakob Kruse, Gianluca Detommaso, Robert Scheichl, and Ullrich Köthe. "HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference". In: *Proceedings of AAAI-2021* (2021). URL: https://arxiv.org/pdf/1905.10687.pdf.

Seismic imaging example setup

$$\mathbf{d} = \mathbf{J}[\mathbf{m}_0, \mathbf{q}]\delta\mathbf{m} + \boldsymbol{\eta} + \mathbf{n}, \quad \mathbf{n} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

- linearized observed data $\mathbf{d}$
- linearized Born demigration operator $\mathbf{J}$
- source signature $\mathbf{q}$
- smooth background model $\mathbf{m}_0$
- unknown perturbation model $\delta\mathbf{m}$
- linearization error and noise $\boldsymbol{\eta}, \mathbf{n}$

training pairs

$$\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^{n} \sim p(\mathbf{y}, \mathbf{x})$$

where

$$(\mathbf{y}_i, \mathbf{x}_i) = \left(\mathbf{J}^\top(\mathbf{J}\delta\mathbf{m}_i + \mathbf{n}_i), \delta\mathbf{m}_i\right)$$

Low- and high-fidelity images

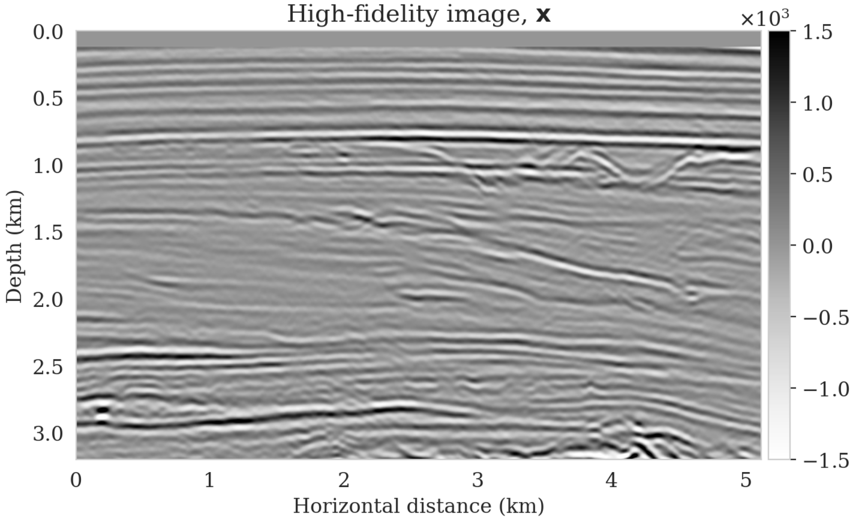
For training, we select $\delta\mathbf{m}$ as

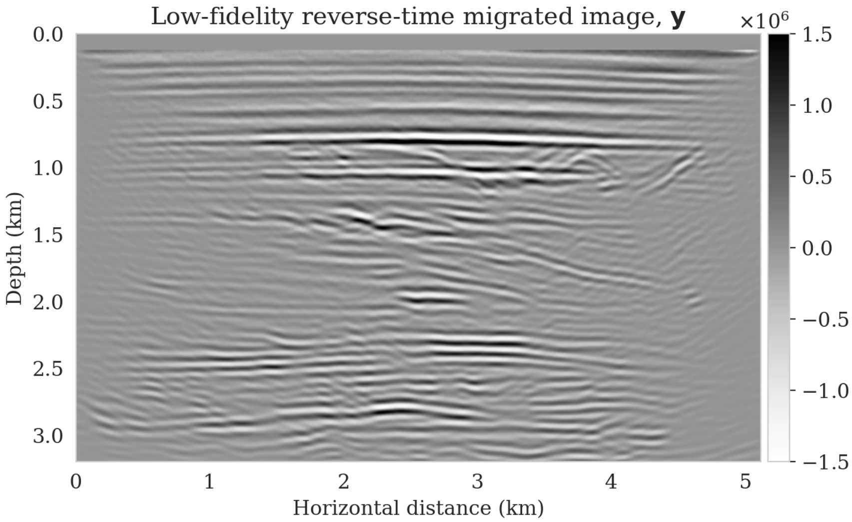
- $3075\,\mathrm{m} \times 5120\,\mathrm{m}$ sections from the Parihaka dataset
- artificial $125\,\mathrm{m}$ water layer to limit the near source imaging artifacts

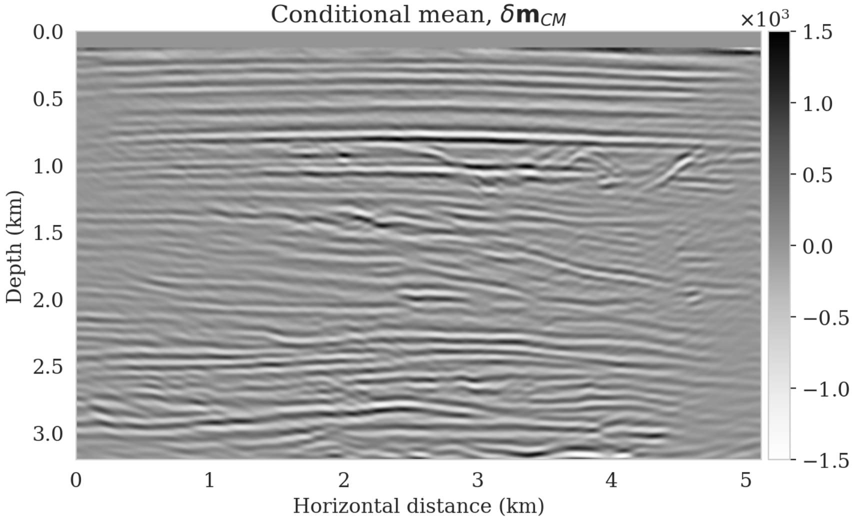
Other possible choices

- 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, ...

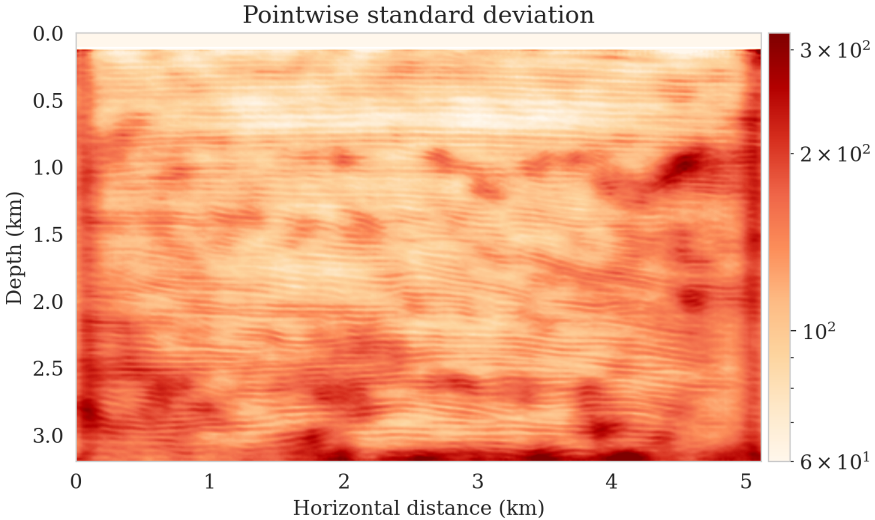
WesternGeco. *Parihaka 3D PSTM Final Processing Report*. Tech. rep. New Zealand Petroleum Report 4582. 2012. URL: https://wiki.seg.org/wiki/Parihaka-3D.


High-fidelity image, **x**
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


Low-fidelity reverse-time migrated image, **y**
×10^6
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5



$$\mathbb{E}\left[\mathbf{x} \mid \mathbf{y}\right] = \int \mathbf{x}\, p_{\text{post}}\left(\mathbf{x} \mid \mathbf{y}\right) \mathrm{d}\mathbf{x}$$



$$\mathbb{E}\left[\,\left(\mathbf{x} - \mathbb{E}\left[\mathbf{x} \mid \mathbf{y}\right]\right)^{\circ 2} \mid \mathbf{y}\right] = \int \left(\mathbf{x} - \mathbb{E}\left[\mathbf{x} \mid \mathbf{y}\right]\right)^{\circ 2} \, p_{\text{post}}\left(\mathbf{x} \mid \mathbf{y}\right) \mathrm{d}\mathbf{x}$$

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

$$T_{\theta_x}(\mathbf{z}) := G^{-1}_{\theta_x}(G_{\theta_y}(\mathbf{y}), \mathbf{z})$$

- preconditioned maximum a posteriori (MAP) estimation

$$\min_{\mathbf{z}} \frac{1}{2\sigma^2} \left\| F\left(T_{\theta_x}(\mathbf{z})\right) - \mathbf{y} \right\|_2^2 + \frac{1}{2} \left\| \mathbf{z} \right\|_2^2$$

- preconditioned physics-based posterior learning

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

Ali Siahkoohi, Gabrio Rizzuti, Mathias Louboutin, Philipp Witte, and Felix J. Herrmann. "Preconditioned training of normalizing flows for variational inference in inverse problems". In: *3rd Symposium on Advances in Approximate Bayesian Inference*. Jan. 2021. URL: https://openreview.net/pdf?id=P9m1sMaNQ8T.

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