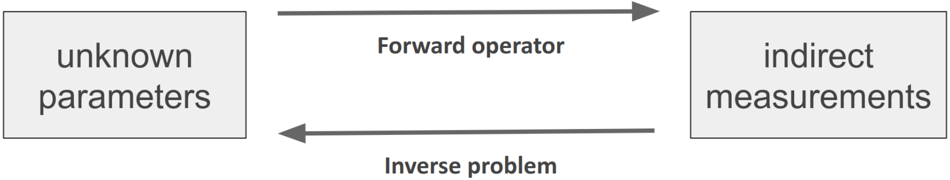
Deep generative models for solving geophysical inverse problems

Ali Siahkoohi

Ph.D. Dissertation Defense

July 7, 2022

SLIM
Georgia Institute of Technology


unknown parameters
Forward operator
Inverse problem
indirect measurements

Bayesian inverse problems

represent the solution as a distribution over the model space

i.e., posterior distribution

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

Choosing a prior distribution

encode prior knowledge

avoid unwanted bias due to overly simplifying priors

Computational cost

costly forward operator

high dimensional sampling/integration

Generative models 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

Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. Proceedings of Machine Learning Research. PMLR, July 2020, pp. 399–409. URL: http://proceedings.mlr.press/v119/asim20a.html.

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.

AmirEhsan Khorashadizadeh, Konik Kothari, Leonardo Salsi, Ali Aghababaei Harandi, Maarten de Hoop, and Ivan Dokmanić. "Conditional Injective Flows for Bayesian Imaging". In: *arXiv preprint arXiv:2204.07664* (2022).

Objective. develop generative-model-based methods that

- enhance the solution quality
- shorten the time to solution
- reduce the cost of Bayesian inference
- are feasible in domains with limited access to training data

Chapter 2. Deep Bayesian inference for seismic imaging with tasks

Ali Siahkoohi [1] Gabrio Rizzuti [2] Felix J. Herrmann [1]

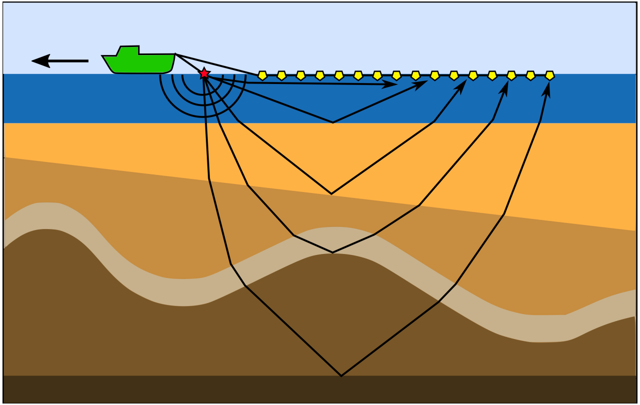
[1]Georgia Institute of Technology

[2]Utrecht University

SLIM
Georgia Institute of Technology

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. "Deep Bayesian inference for seismic imaging with tasks". In: *Geophysics* 87.5 (June 2022). DOI: 10.1190/geo2021-0666.1. URL: https://arxiv.org/abs/2110.04825.

Seismic imaging



Philipp A. Witte. "Software and algorithms for large-scale seismic inverse problems". PhD thesis. Georgia Institute of Technology, Feb. 2020. URL: https://slim.gatech.edu/Publications/Public/Thesis/2020/witte2020THsal/witte2020THsal.pdf.

Forward model

$$\mathbf{d}_i = \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m}^* + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$

linearized Born scattering operator $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$

smooth squared-slowness model of the subsurface $\mathbf{m}_0$

seismic source time signature $\mathbf{q}_i$

observed data $\mathbf{d} = \{\mathbf{d}_i\}_{i=1}^{n_s}$

Linearized Born modeling operator

$$\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i) = \nabla_{\mathbf{m}} \, \mathbf{P}\mathbf{A}^{-1}[\mathbf{m}]\mathbf{q}_i \Big|_{\mathbf{m}=\mathbf{m}_0}$$

$$= -\mathbf{P} \underbrace{\mathbf{A}^{-1}[\mathbf{m}_0] \, \mathrm{diag}\left(\nabla \mathbf{A}[\mathbf{m}_0] \overbrace{\mathbf{A}^{-1}[\mathbf{m}_0]\mathbf{q}_i}^{\text{PDE solve}}\right)}_{\text{PDE solve}}$$

discretized acoustic wave equation $\mathbf{A}[\mathbf{m}]$

receiver restriction linear operator $\mathbf{P}$

Proposed method

regularization via the inductive bias of network architecture

posterior sampling via stochastic gradient Langevin dynamics

translating uncertainties into downstream tasks

V. Lempitsky, A. Vedaldi, and D. Ulyanov. "Deep Image Prior". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 9446–9454. DOI: 10.1109/CVPR.2018.00984.

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

$$-\log p_{\text{post}}\left(\mathbf{w} \mid \mathbf{d}\right) = \underbrace{\frac{1}{2\sigma^2}\sum_{i=1}^{n_s}\|\mathbf{d}_i - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)g(\mathbf{z}, \mathbf{w})\|_2^2}_{\text{negative-log likelihood}} + \underbrace{\frac{\lambda^2}{2}\|\mathbf{w}\|_2^2}_{\substack{\text{negative-log}\\\text{prior}}} + \underbrace{\text{const}}_{\text{Ind. of } \mathbf{w}}$$

untrained convolutional neural network (CNN) $g(\mathbf{z}, \mathbf{w})$

CNN weights $\mathbf{w}$

fixed input $\mathbf{z}$

Stochastic gradient Langevin dynamics (SGLD)

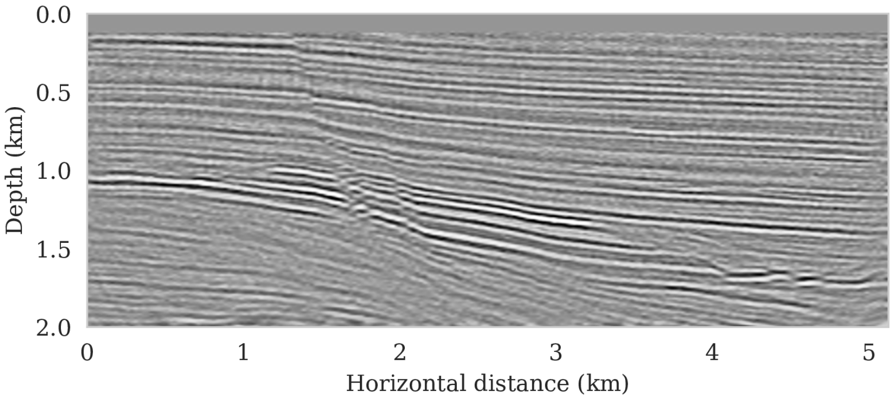
$$\mathbf{w}_{k+1} = \mathbf{w}_k - \frac{\alpha_k}{2}\mathbf{M}_k \nabla_{\mathbf{w}} \log p_{\text{post}}\left(\mathbf{w} \mid \mathbf{d}\right) + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathrm{N}(\mathbf{0}, \alpha_k \mathbf{M}_k)$$

preconditioning matrix $\mathbf{M}_k$

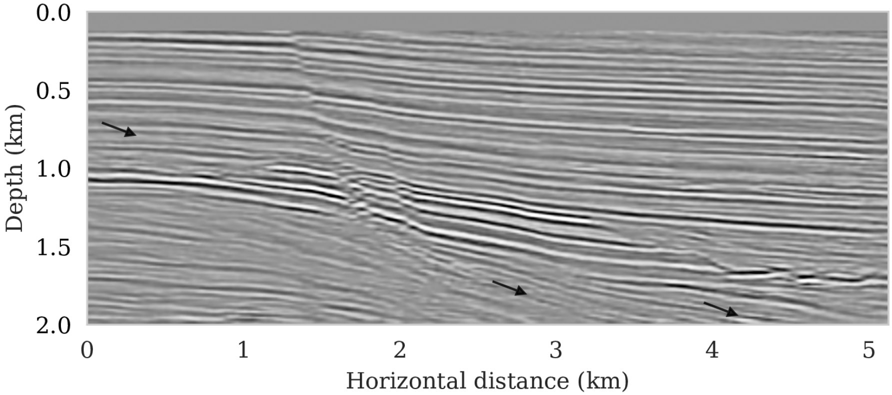
stepsize α_k

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

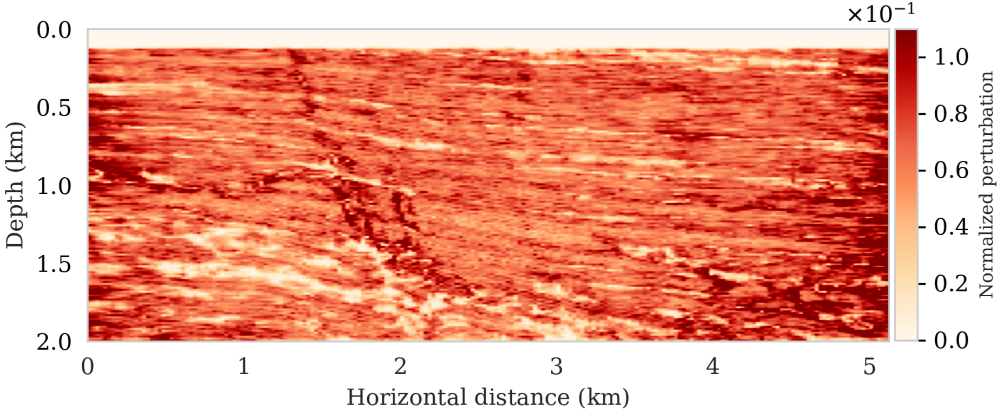
Chunyuan Li, Changyou Chen, David Carlson, and Lawrence Carin. "Preconditioned Stochastic Gradient Langevin Dynamics for Deep Neural Networks". In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. Phoenix, Arizona: AAAI Press, 2016, pp. 1788–1794. DOI: 10.5555/3016100.3016149. URL: https://dl.acm.org/doi/abs/10.5555/3016100.3016149.



estimation with no prior—SNR: 8.25 dB



conditional mean estimate—SNR: 9.66 dB



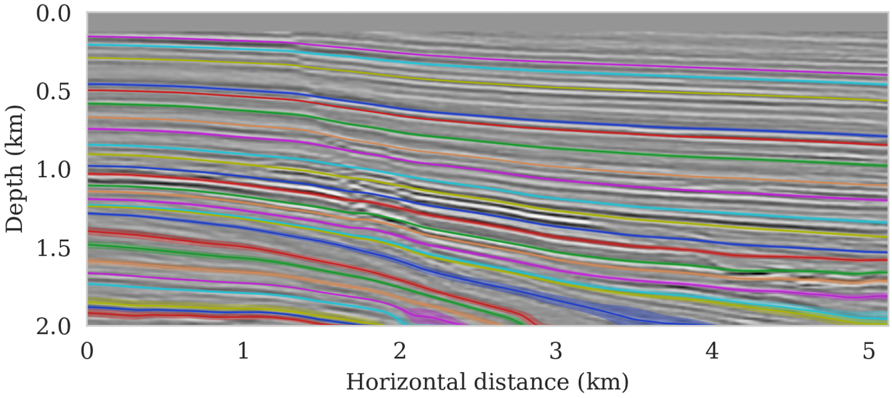
normalized pointwise standard deviation

Downstream task inference

$$\mathbb{E}_{\mathbf{h}\sim p_{\text{post}}(\mathbf{h}|\mathbf{d})}\left[f\left(\mathbf{h}\right)\right] = \underbrace{\mathbb{E}_{\delta\mathbf{m}\sim p_{\text{post}}(\delta\mathbf{m}|\mathbf{d})} \underbrace{\mathbb{E}_{\mathbf{h}\sim p(\mathbf{h}|\delta\mathbf{m})}\left[f\left(\mathbf{h}\right)\right].}_{\substack{\text{uncertainty in} \\ \text{horizon tracking}}}}_{\text{uncertainty in seismic imaging}}$$

outcome (random variable) of downstream task $\mathbf{h}$

assuming $\quad \mathbf{h} \perp\!\!\!\perp \delta\mathbf{d} \mid \delta\mathbf{m} \quad \Leftrightarrow \quad p\left(\mathbf{h} \mid \delta\mathbf{m}, \delta\mathbf{d}\right) = p\left(\mathbf{h} \mid \delta\mathbf{m}\right)$



Uncertainty in horizon tracking

Chapter 3. Weak deep priors for seismic imaging

Ali Siahkoohi [1] Gabrio Rizzuti [2] Felix J. Herrmann [1]

[1]Georgia Institute of Technology

[2]Utrecht University

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. "Weak deep priors for seismic imaging". In: *90th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Sept. 2020, pp. 2998–3002. DOI: 10.1190/segam2020-3417568.1. URL: https://arxiv.org/abs/2004.06835.

Contribution

weak deep priors—a computationally practical formulation

relaxing the deep prior—i.e., $\delta\mathbf{m} \sim \mathrm{N}(g(\mathbf{z}, \mathbf{w}), \gamma^{-2}\mathbf{I})$

$$\min_{\delta\mathbf{m},\,\mathbf{w}} \left[\frac{1}{2\sigma^2} \sum_{i=1}^{n_s} \|\delta\mathbf{d}_i - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m}\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{w}\|_2^2 \right]$$

subject to $\delta\mathbf{m} = g(\mathbf{z}, \mathbf{w})$

(deep prior)

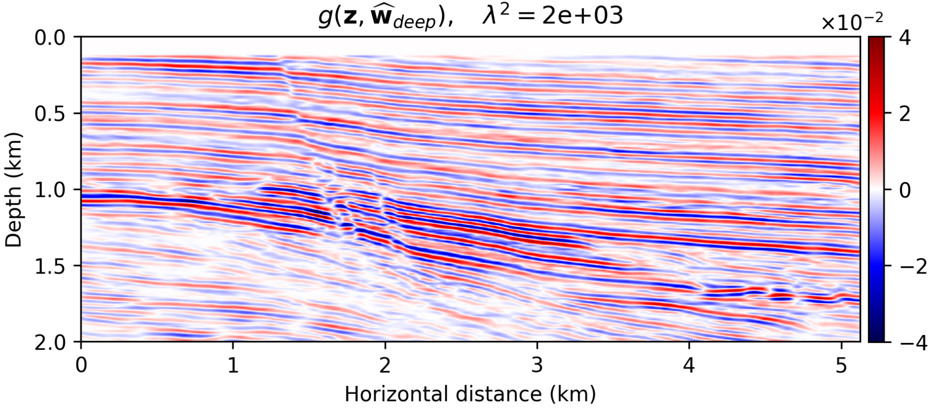
subject to $\delta\mathbf{m} = g(\mathbf{z}, \mathbf{w}) + \mathrm{N}(\mathbf{0}, \gamma^{-2}\mathbf{I})$

(weak deep prior)

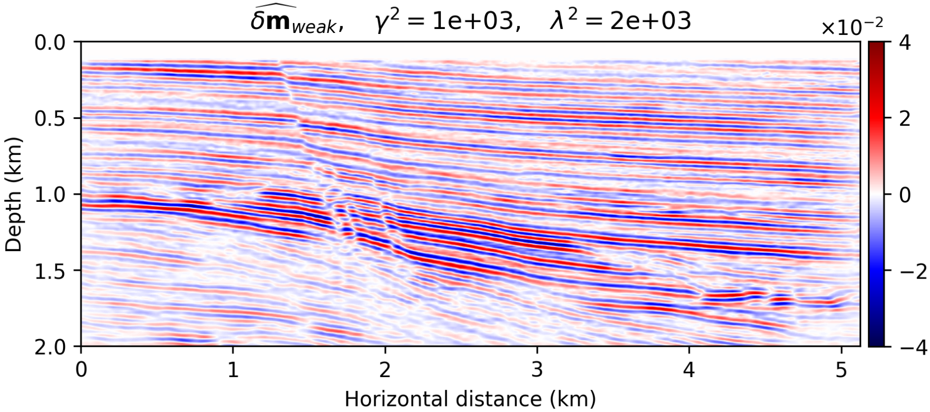
Weak deep-prior based imaging

$$\min_{\delta\mathbf{m},\,\mathbf{w}} \quad \underbrace{\frac{1}{2\sigma^2}\sum_{i=1}^{n_s}\|\delta\mathbf{d}_i - \mathbf{J}(\mathbf{m}_0,\mathbf{q}_i)\delta\mathbf{m}\|_2^2}_{\text{negative-log likelihood}} + \underbrace{\frac{\gamma^2}{2}\|\delta\mathbf{m} - \mathbf{g}(\mathbf{z},\mathbf{w})\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{w}\|_2^2}_{\text{negative-log (weak) prior}}$$

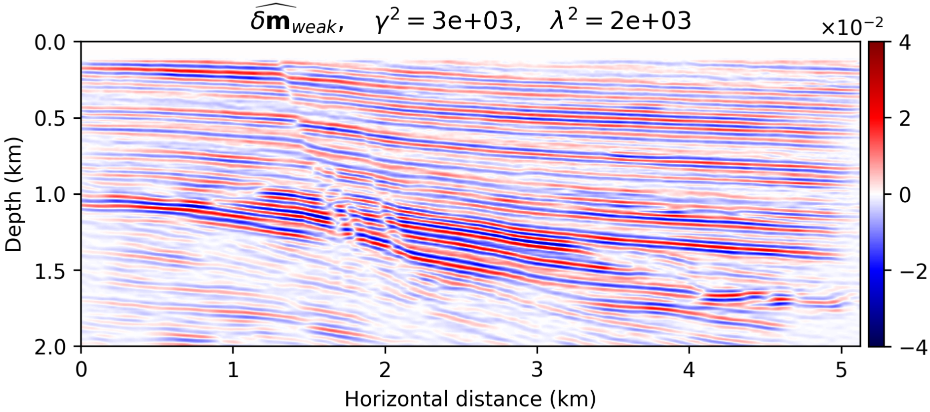
$g(\mathbf{z}, \widehat{\mathbf{w}}_{deep}), \quad \lambda^2 = 2\text{e}+03$



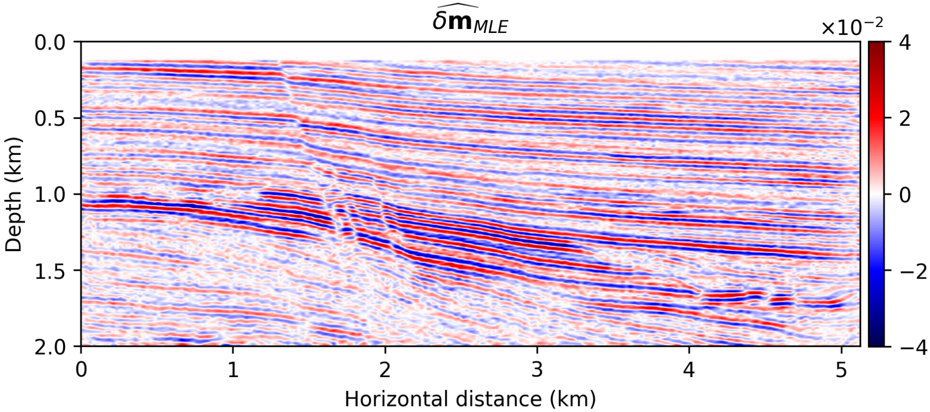
imaging with deep prior, cost $\approx$ 15 reverse-time migrations



imaging with weak deep prior, cost $\approx$ 2 reverse-time migrations



imaging with weak deep prior, cost $\approx$ 2 reverse-time migrations



MLE — no regularization (prior)

Bayesian inference via weak deep priors?

expectation maxzimization (EM) + SGLD in the latent space

Felix J. Herrmann, Ali Siahkoohi, and Gabrio Rizzuti. "Learned imaging with constraints and uncertainty quantification". In: *Neural Information Processing Systems (NeurIPS) 2019 Deep Inverse Workshop*. Dec. 2019. URL: `https://arxiv.org/pdf/1909.06473.pdf`.

Chapter 4. Preconditioned training of normalizing flows for variational inference in inverse problems

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

SLIM
Georgia Institute of Technology

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.

Markov chain Monte Carlo	Variational inference
asymptotically exact samples	reduces sampling to optimization
high-dimensional integration/sampling	trading off computational cost for accuracy
costs of forward operator	cheap sampling after optimization
requires choosing a prior distribution	requires choosing a prior distribution

Amortized variational inference

$$\phi^* = \underset{\phi}{\arg\min}\ \mathbb{E}_{\mathbf{y}\sim p_{\text{data}}(\mathbf{y})}\left[\mathbb{KL}\left(p_{\text{post}}(\mathbf{x}\mid\mathbf{y})\,||\,p_\phi(\mathbf{x}\mid\mathbf{y})\right)\right]$$

forward model $\mathbf{y} = \mathcal{F}(\mathbf{x}) + \boldsymbol{\epsilon}$

computationally costly forward operator $\mathcal{F}:\mathcal{X}\to\mathcal{Y}$

surrogate conditional distribution $p_\phi(\mathbf{x}\mid\mathbf{y})$

Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. "BayesFlow: Learning complex stochastic models with invertible neural networks". In: *IEEE transactions on neural networks and learning systems* (2020).

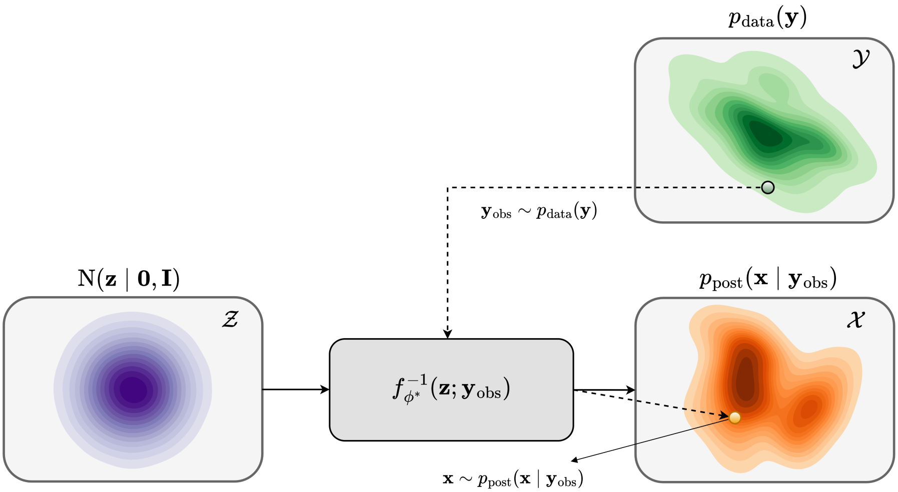
Normalizing flows for amortized inference

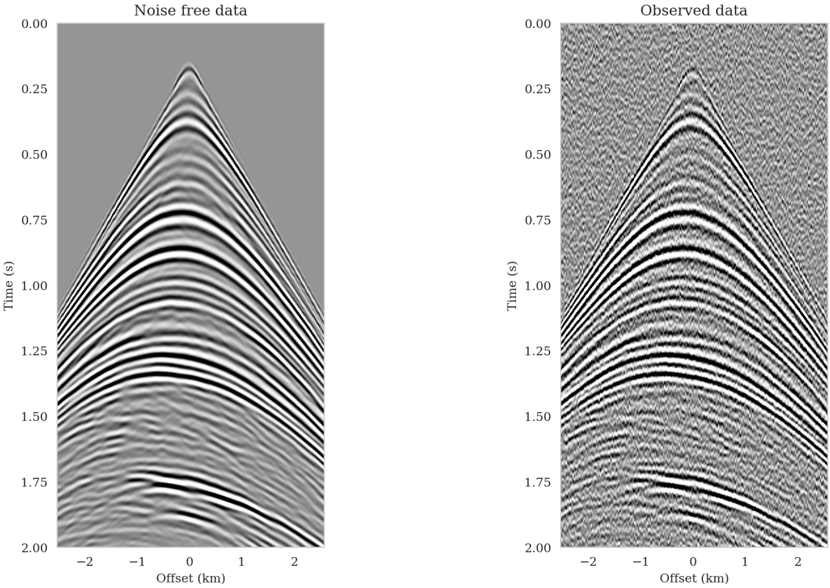
$$\phi^* = \underset{\phi}{\arg\min}\ \mathbb{E}_{\mathbf{x},\mathbf{y}\sim p(\mathbf{x},\mathbf{y})}\left[\underbrace{\frac{1}{2}\left\|f_\phi(\mathbf{x};\mathbf{y})\right\|_2^2}_{\text{normalizes the input}} - \underbrace{\log\left|\det \nabla_{\mathbf{x}} f_\phi(\mathbf{x};\mathbf{y})\right|}_{\substack{\text{entropy regularization} \\ \text{e.g., avoids } f_\phi(\mathbf{x};\mathbf{y}) \equiv \mathbf{0}}} \right]$$

$f_\phi(\,\cdot\,;\mathbf{y}) : \mathcal{X} \to \mathcal{Z}$ an invertible neural net

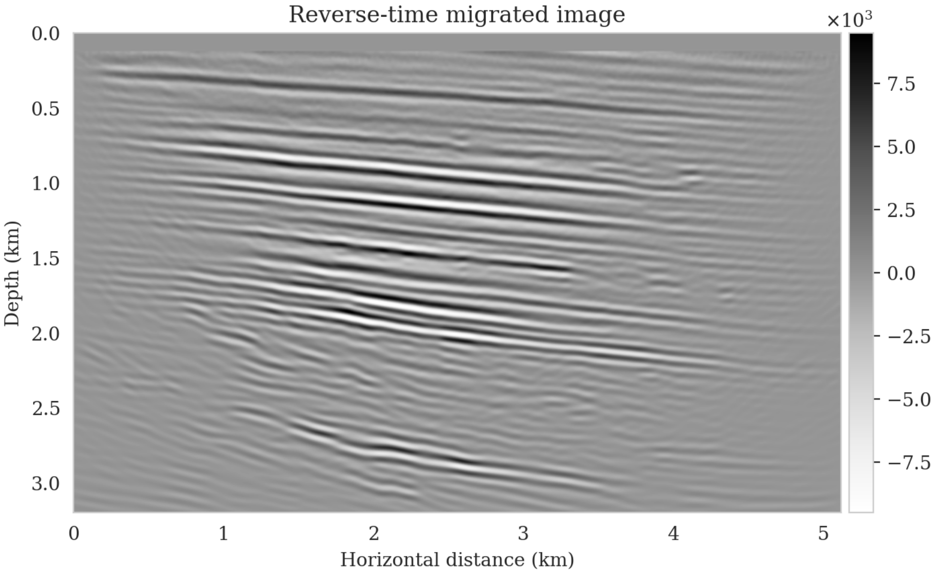
negligible computational cost of $\det \nabla_{\mathbf{x}} f_\phi(\mathbf{x};\mathbf{y})$'s gradient due to f_ϕ's architecture

Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. "BayesFlow: Learning complex stochastic models with invertible neural networks". In: *IEEE transactions on neural networks and learning systems* (2020).

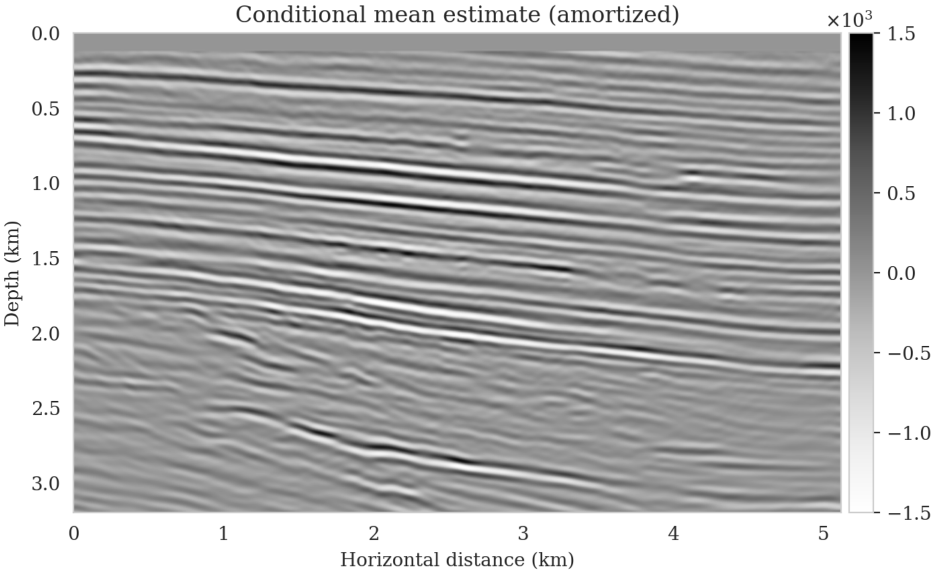

$p_{\text{data}}(\mathbf{y})$
$\mathcal{Y}$
$\mathbf{y}_{\text{obs}} \sim p_{\text{data}}(\mathbf{y})$
$\mathrm{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$
$\mathcal{Z}$
$f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}})$
$p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$
$\mathcal{X}$
$\mathbf{x} \sim p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$



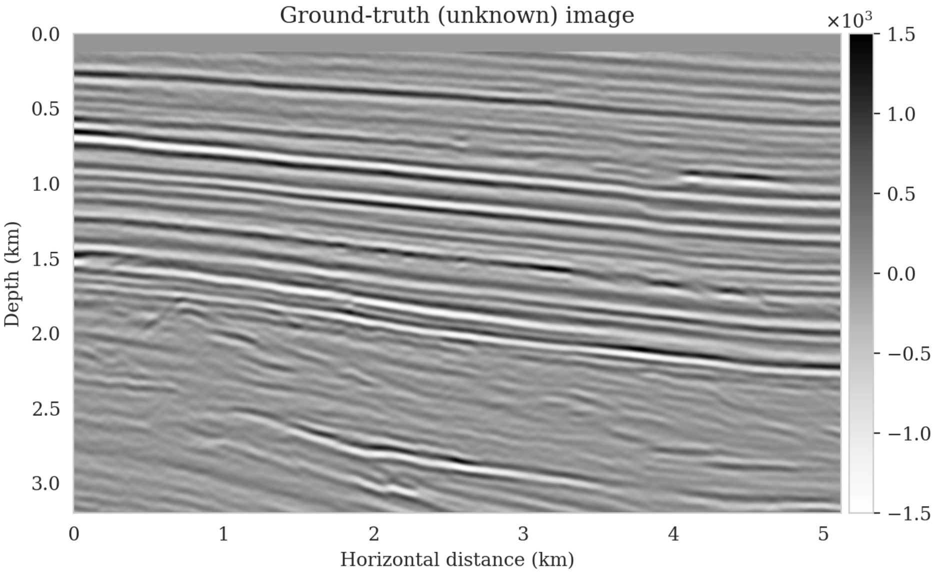
(left) noise free data (right) noisy data


Reverse-time migrated image
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10^3
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

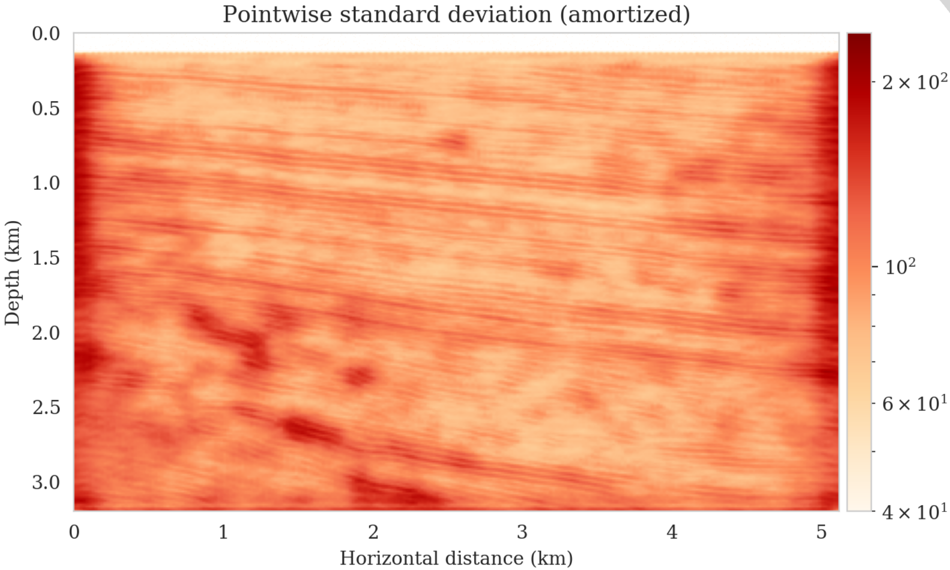
SNR −12.17 dB


Conditional mean estimate (amortized)
Depth (km)
Horizontal distance (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 11.24 dB



previously unseen (test) seismic image


Pointwise standard deviation (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
2×10^2
10^2
6×10^1
4×10^1

Multi-fidelity preconditioned scheme

exploit the learned information during amortized variational inference

ensure physics and data fidelity

Finetune the amortized posterior approximation

$$\min_{\phi^*} \mathbb{KL}\left(p_{\phi^*}(\mathbf{x} \mid \mathbf{y}_{\text{obs}}) \,\big|\big|\, p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}})\right)$$

mitigating errors to data distribution shifts $\mathbf{y}_{\text{obs}} \sim \widehat{p}_{\text{data}}(\mathbf{y})$

finetuning ϕ^* to better approximate the posterior

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.

Multi-fidelity preconditioned scheme

$$\min_{\phi^*} \mathbb{E}_{\mathbf{z}\sim \mathrm{N}(\mathbf{0},\mathbf{I})} \Bigg[\frac{1}{2\sigma^2} \left\| \mathbf{y}_{\text{obs}} - \mathcal{F} \circ f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}}) \right\|_2^2$$

$$- \log p_{\text{prior}}\left(f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}})\right) - \log \left| \det \nabla_{\mathbf{z}} f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}}) \right| \Bigg]$$

pretrained $f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}})$ already approximates the posterior

less costly evaluations of $\mathcal{F}$ than non-preconditioned approach

learned conditional prior $p_{\text{prior}}(\mathbf{x}) := \mathrm{N}\left(f_{\phi^*}(\mathbf{x}; \mathbf{y}_{\text{obs}}) \mid \mathbf{0}, \mathbf{I}\right) \left| \det \nabla_{\mathbf{x}} f_{\phi^*}(\mathbf{x}; \mathbf{y}_{\text{obs}}) \right|$

Chapter 5. Reliable amortized variational inference via physics-based latent distribution correction

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

[1]Georgia Institute of Technology

[2]Utrecht University

SLIM

Georgia Institute of Technology

Ali Siahkoohi, Gabrio Rizzuti, Rafael Orozco, and Felix J. Herrmann. "Reliable amortized variational inference via physics-based latent distribution correction". In: *To be submitted to Geophysics* (July 2022).

Implications of imperfect pretraining or data distribution shifts

inaccurate normalization through the conditional normalizing flow

$$p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}}) \neq \mathrm{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$$

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

Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. "BayesFlow: Learning complex stochastic models with invertible neural networks". In: *IEEE transactions on neural networks and learning systems* (2020).

Implications of imperfect pretraining or data distribution shifts

inaccurate posterior sampling given $\mathbf{z} \sim \mathrm{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$ as input

$$p_\phi(\mathbf{x} \mid \mathbf{y}_{\text{obs}}) \neq p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$$

$p_\phi(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$ predicted posterior distribution

Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. "BayesFlow: Learning complex stochastic models with invertible neural networks". In: *IEEE transactions on neural networks and learning systems* (2020).

SLIM

Latent distribution correction

$$\mathbf{z} \sim \mathrm{N}\big(\mathbf{z} \mid \mathbf{0}, \mathbf{I}\big) \quad \longrightarrow \quad \mathbf{z} \sim \mathrm{N}\big(\mathbf{z} \mid \boldsymbol{\mu}, \mathrm{diag}(\mathbf{s})^2\big)$$

$p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}})$ not too far from normal distribution

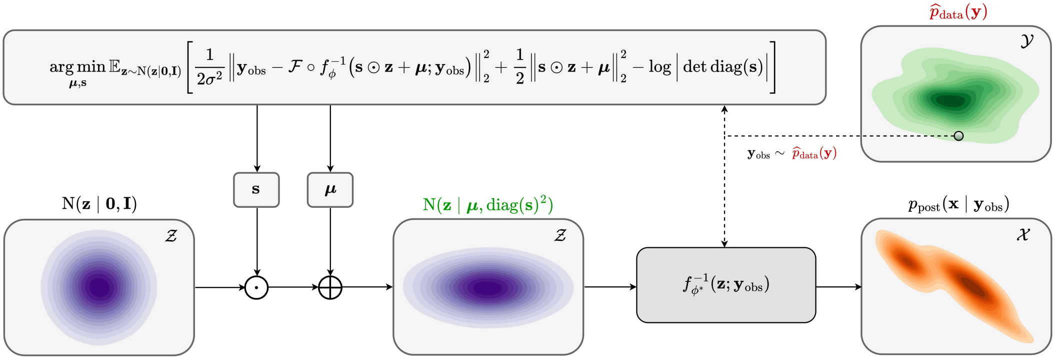
relaxing the initial approximation $p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}}) \approx \mathrm{N}\big(\mathbf{z} \mid \mathbf{0}, \mathbf{I}\big)$

Physics-based latent distribution correction

$$\min_{\boldsymbol{\mu},\mathbf{s}} \mathbb{KL}\left(\mathrm{N}(\mathbf{z} \mid \boldsymbol{\mu}, \mathrm{diag}(\mathbf{s})^2) \,||\, p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}})\right), \quad \mathbf{y}_{\text{obs}} \sim \widehat{p}_{\text{data}}(\mathbf{y})$$

with

$$-\log p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}}) = \frac{1}{2\sigma^2}\left\|\mathbf{y}_{\text{obs}} - \mathcal{F} \circ f_\phi(\mathbf{z}; \mathbf{y}_{\text{obs}})\right\|_2^2 + \frac{1}{2}\left\|\mathbf{z}\right\|_2^2 + \text{const.}$$



$\arg\min_{\boldsymbol{\mu},\mathbf{s}} \mathbb{E}_{\mathbf{z}\sim \mathrm{N}(\mathbf{z}|\mathbf{0},\mathbf{I})} \left[\frac{1}{2\sigma^2} \left\| \mathbf{y}_{\text{obs}} - \mathcal{F} \circ f_{\phi}^{-1}(\mathbf{s} \odot \mathbf{z} + \boldsymbol{\mu}; \mathbf{y}_{\text{obs}}) \right\|_2^2 + \frac{1}{2} \left\| \mathbf{s} \odot \mathbf{z} + \boldsymbol{\mu} \right\|_2^2 - \log \left| \det \operatorname{diag}(\mathbf{s}) \right| \right]$
$\widehat{p}_{\text{data}}(\mathbf{y})$
$\mathcal{Y}$
$\mathbf{y}_{\text{obs}} \sim \widehat{p}_{\text{data}}(\mathbf{y})$
$\mathbf{s}$
$\boldsymbol{\mu}$
$\mathrm{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$
$\mathcal{Z}$
$\mathrm{N}(\mathbf{z} \mid \boldsymbol{\mu}, \operatorname{diag}(\mathbf{s})^2)$
$\mathcal{Z}$
$f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}})$
$p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$
$\mathcal{X}$

$$\min_{\boldsymbol{\mu},\mathbf{s}} \mathbb{E}_{\mathbf{z}\sim \mathrm{N}(\mathbf{0},\mathbf{I})} \Bigg[\frac{1}{2\sigma^2} \sum_{i=1}^{N} \Big\| \mathbf{y}_{\mathrm{obs},i} - \mathcal{F}_i \circ f_{\phi}^{-1}\big(\mathbf{s}\odot\mathbf{z}+\boldsymbol{\mu};\mathbf{y}_{\mathrm{obs}}\big) \Big\|_2^2 + \frac{1}{2} \Big\| \mathbf{s}\odot\mathbf{z}+\boldsymbol{\mu} \Big\|_2^2 - \log \Big| \det \operatorname{diag}(\mathbf{s}) \Big| \Bigg]$$

initializing with $\boldsymbol{\mu} = \mathbf{0}$ and $\operatorname{diag}(\mathbf{s})^2 = \mathbf{I}$

initialization acts as a warm-start and an implicit regularization

expected to be solved relatively cheaply due to the amortization of f_{ϕ}

non-amortized, i.e., specific to one set of observations $\mathbf{y}_{\mathrm{obs}} \sim \widehat{p}_{\mathrm{data}}(\mathbf{y})$

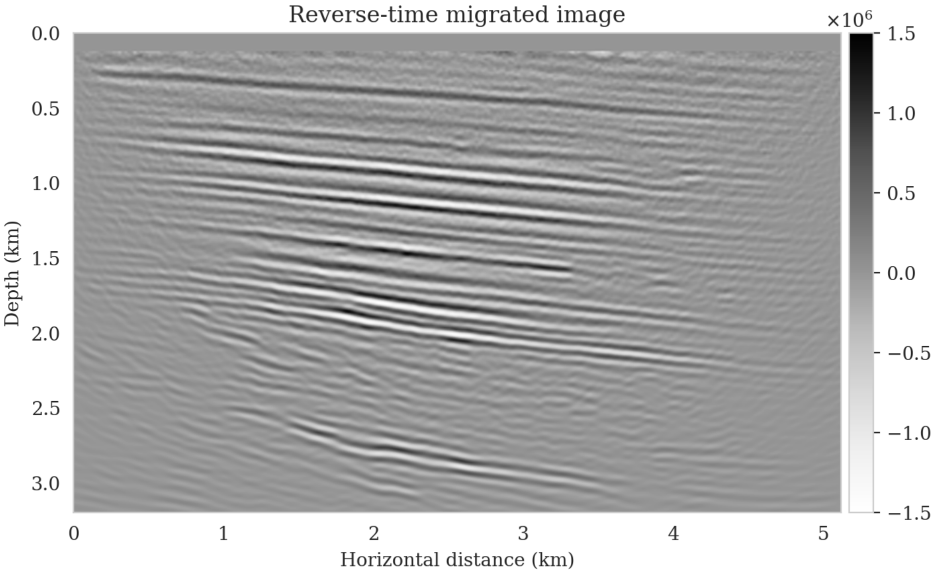
Introducing distribution shifts

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

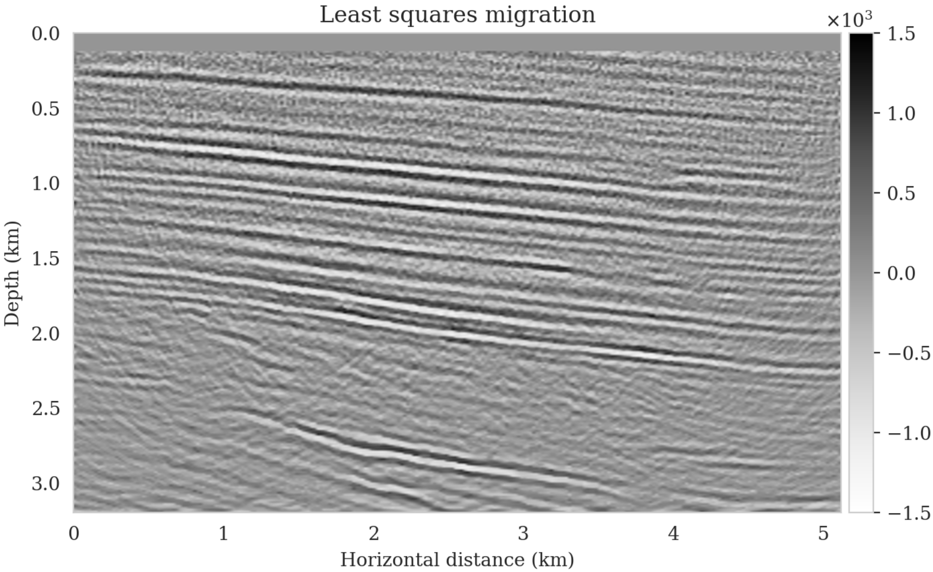
$4\times$ less sources

Physics-based latent distribution correction

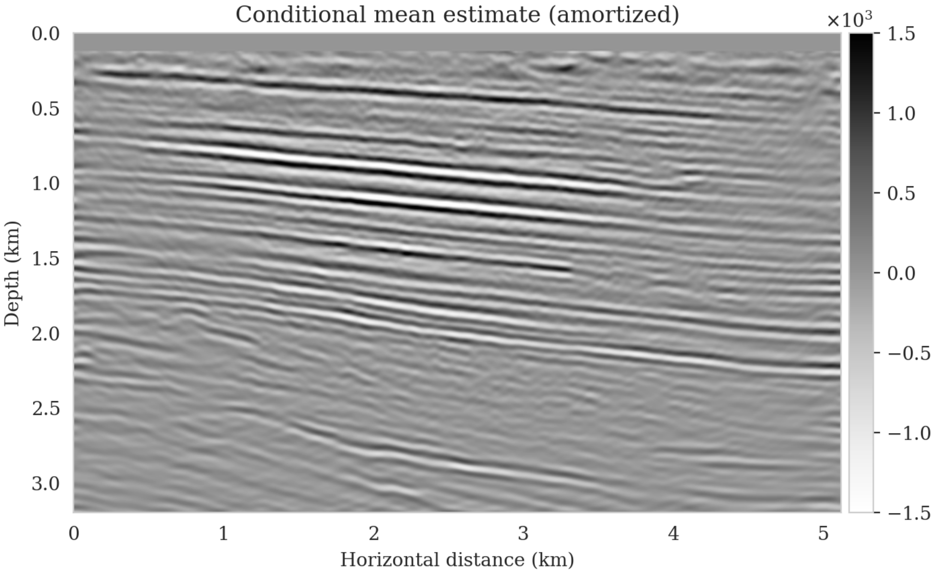
computational cost: approximately $5\times$ RTMs



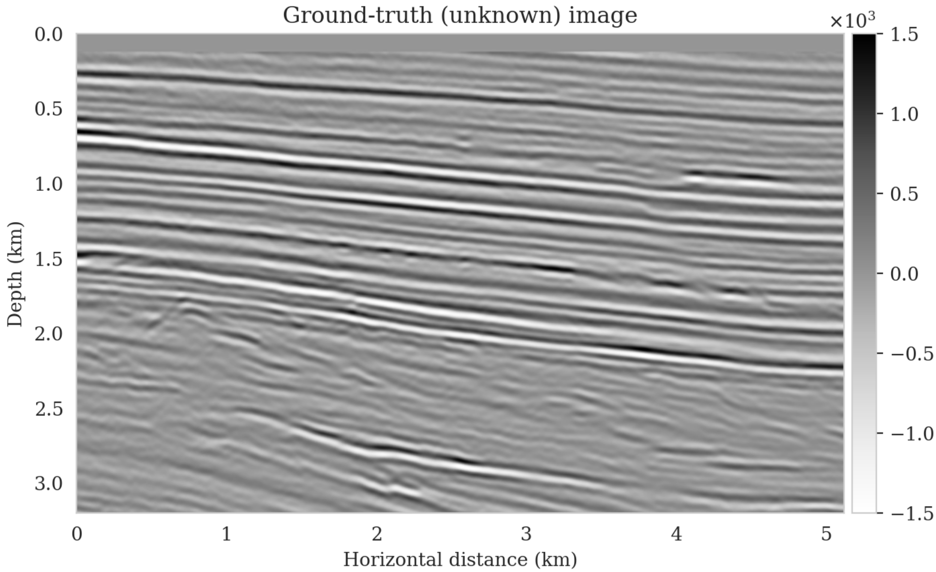
SNR $-8.22\,\mathrm{dB}$


Least squares migration
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

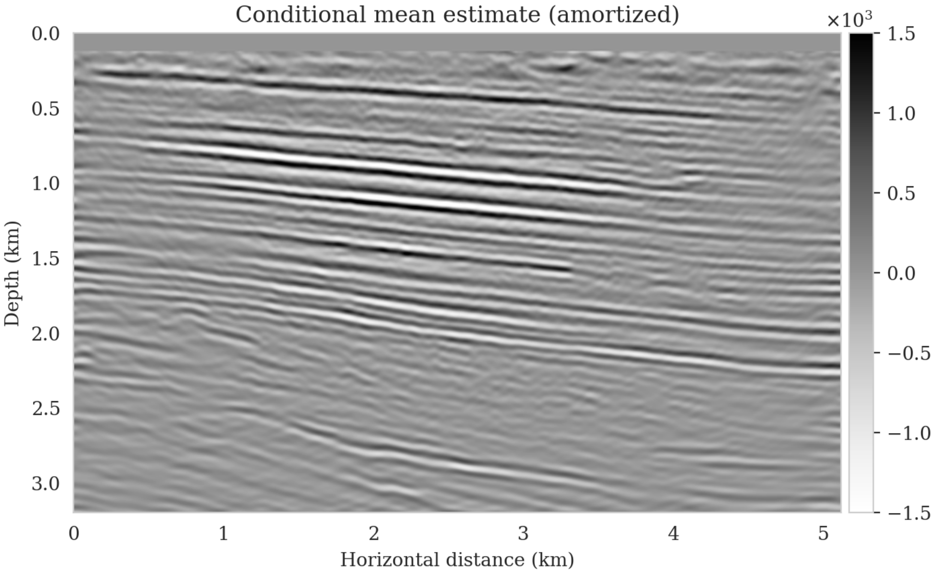
SNR 6.90 dB


Conditional mean estimate (amortized)
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)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

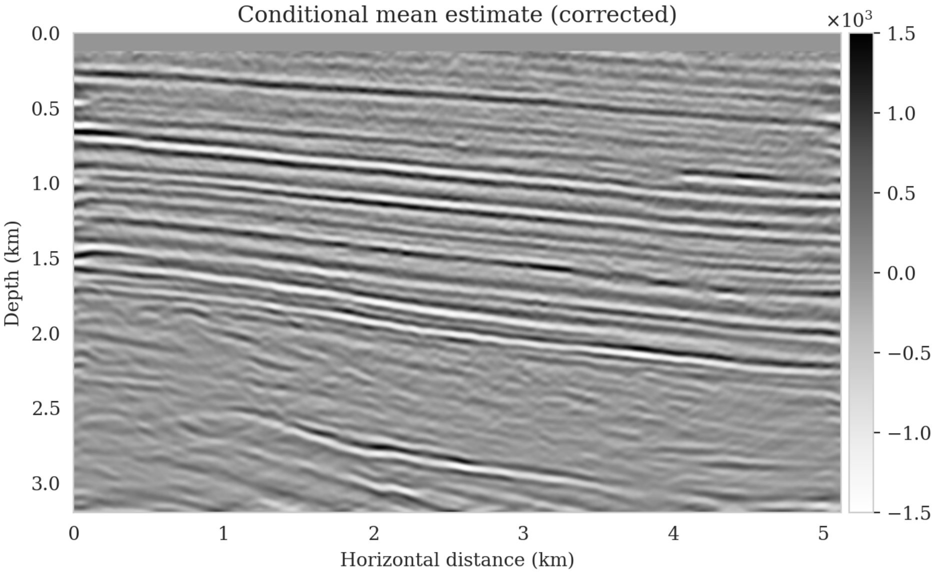
SNR 6.29 dB



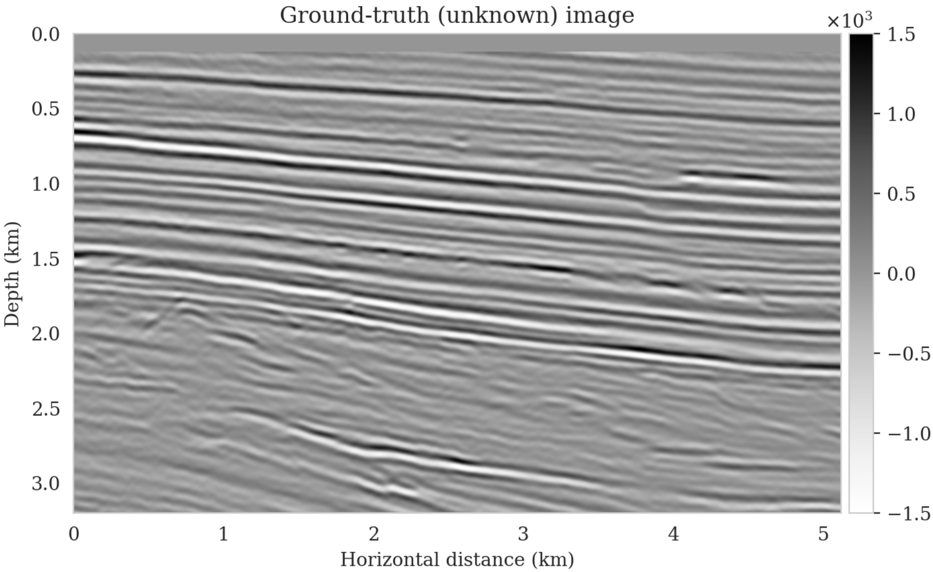
previously unseen (test) seismic image


Conditional mean estimate (amortized)
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)
$\times 10^3$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

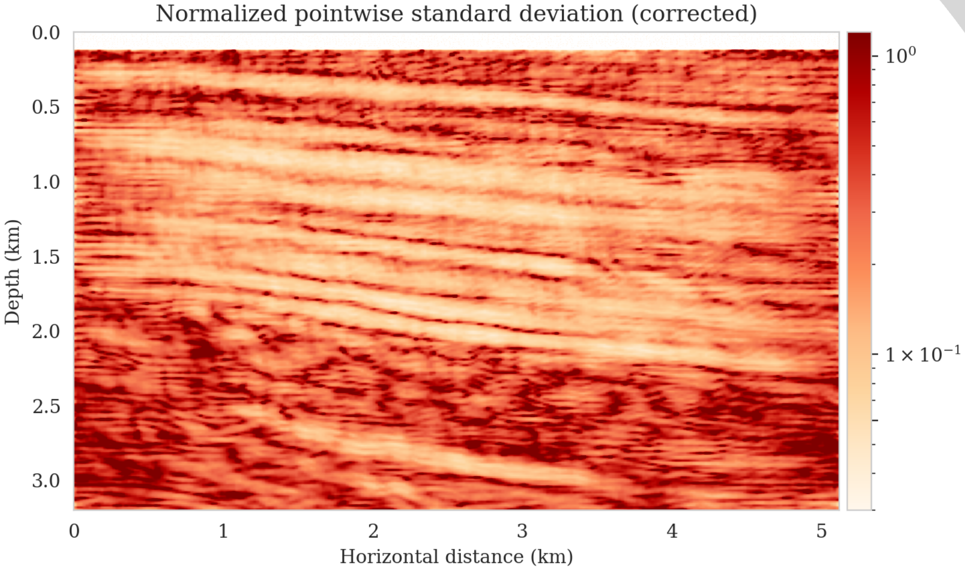
SNR 6.29 dB


Conditional mean estimate (corrected)
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

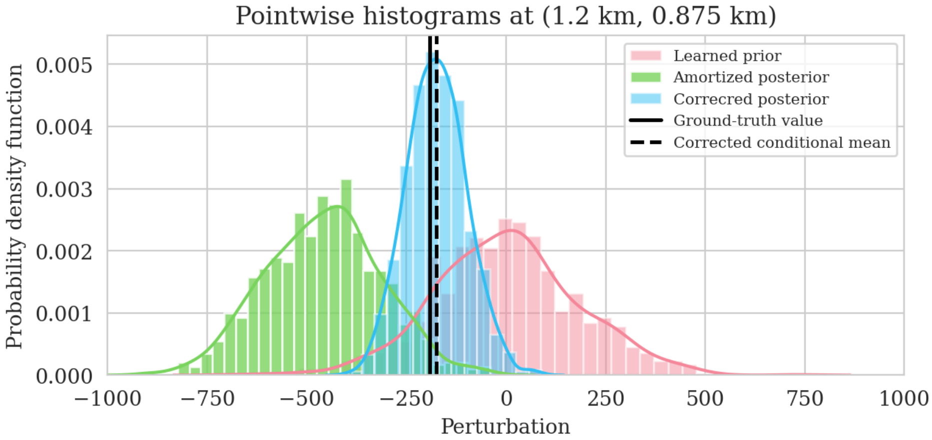
SNR 10.36 dB



previously unseen (test) seismic image

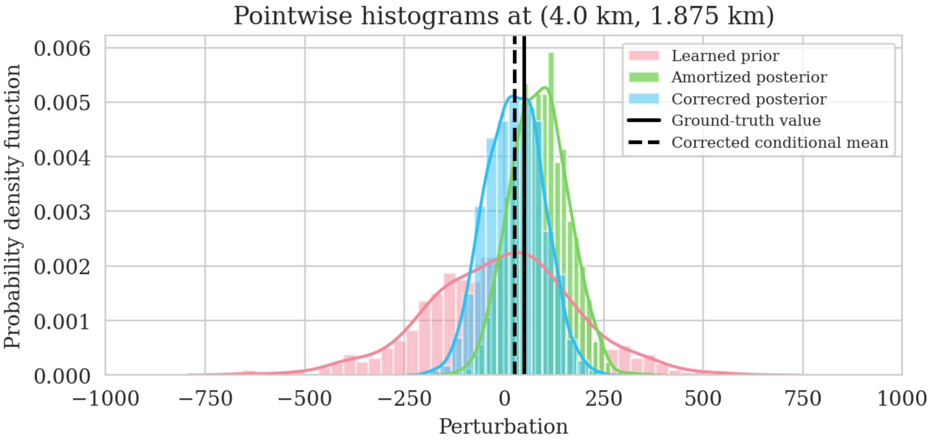


normalized by the the envelope of the conditional mean

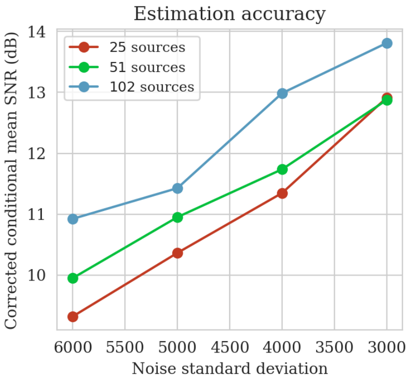


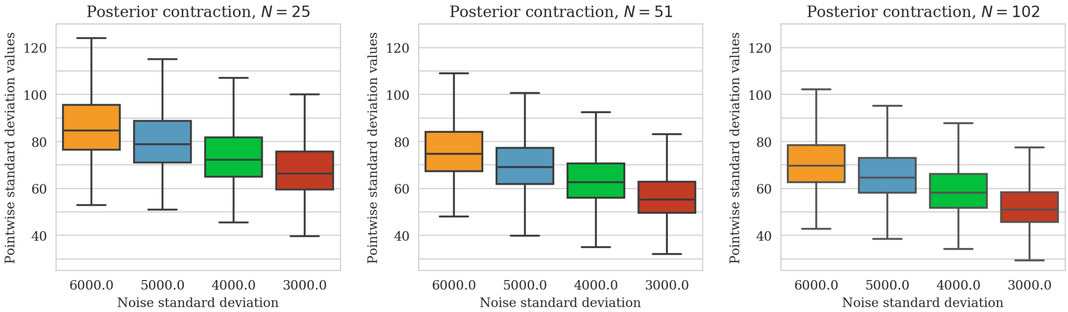
Pointwise histograms at (1.2 km, 0.875 km)
Probability density function
0.005
0.004
0.003
0.002
0.001
0.000
−1000
−750
−500
−250
0
250
500
750
1000
Perturbation
Learned prior
Amortized posterior
Correcred posterior
Ground-truth value
Corrected conditional mean

SLIM



Pointwise histograms at (4.0 km, 1.875 km)
Probability density function
0.006
0.005
0.004
0.003
0.002
0.001
0.000
−1000
−750
−500
−250
0
250
500
750
1000
Perturbation
Learned prior
Amortized posterior
Correcred posterior
Ground-truth value
Corrected conditional mean


Estimation accuracy
25 sources
51 sources
102 sources
14
13
12
11
10
6000
5500
5000
4500
4000
3500
3000
Corrected conditional mean SNR (dB)
Noise standard deviation



(left) $N = 25$, (middle) $N = 51$, and (right) $N = 102$.

Chapter 6. The importance of transfer learning in seismic modeling and imaging

Ali Siahkoohi [1] Mathias Louboutin [1] Felix J. Herrmann [1]

[1]Georgia Institute of Technology

SLIM
Georgia Institute of Technology

Ali Siahkoohi, Mathias Louboutin, and Felix J. Herrmann. "The importance of transfer learning in seismic modeling and imaging". In: *Geophysics* 84.6 (Nov. 2019), A47–A52. DOI: 10.1190/geo2019-0056.1.

Find $\delta\mathbf{m}$ such that

$$\mathbf{d}_i = \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} + \boldsymbol{\epsilon}_i$$

linearized Born scattering operator $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$

measurement noise $\boldsymbol{\epsilon}_i$

Find $\delta\mathbf{m}$ such that

$$\mathbf{d}_i = \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} + \boldsymbol{\epsilon}_i + \boldsymbol{\eta}_i$$

linearized Born scattering operator $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$

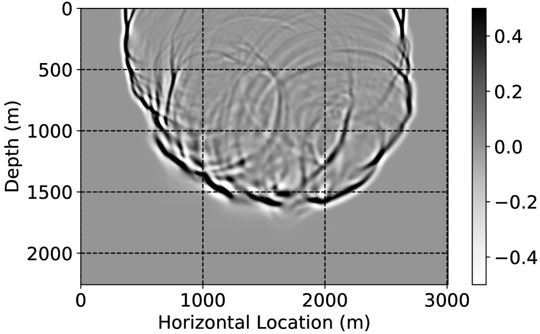
measurement noise $\boldsymbol{\epsilon}_i$ and modeling errors $\boldsymbol{\eta}_i$

Modeling errors include

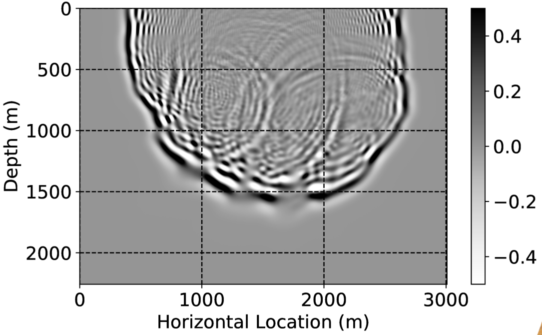
errors in numerical simulations, e.g., coarse finite-difference discretizations

errors in forward model parameterization, e.g., wrong background model $\mathbf{m}_0$ in $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$

Fine spatial discretization
computationally intensive



Coarse spatial discretization
suffers from numerical dispersion



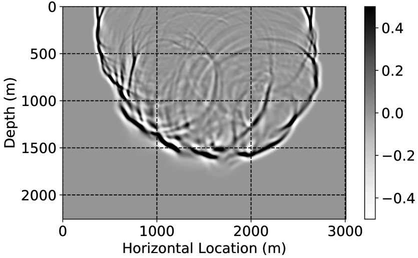
Contributions

train a network that corrects for numerical dispersion

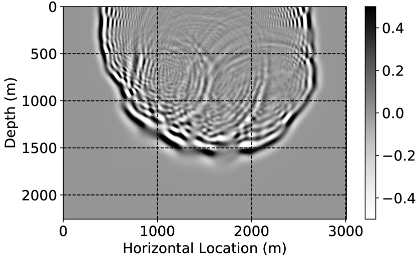
learned learning misfit via GANs to minimize the bias of hand-chosen losses

pretrain with existing low- and high-fidelity pairs, e.g., from neighboring surveys

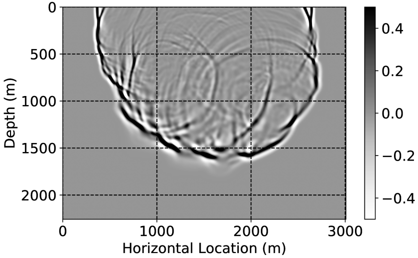
finetune with few low- and high-fidelity pairs from the pertinent survey



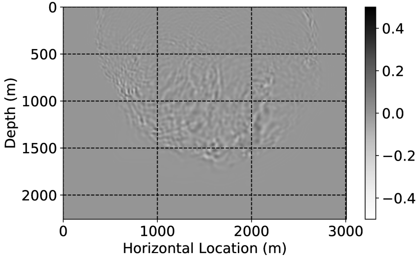
high-fidelity simulation



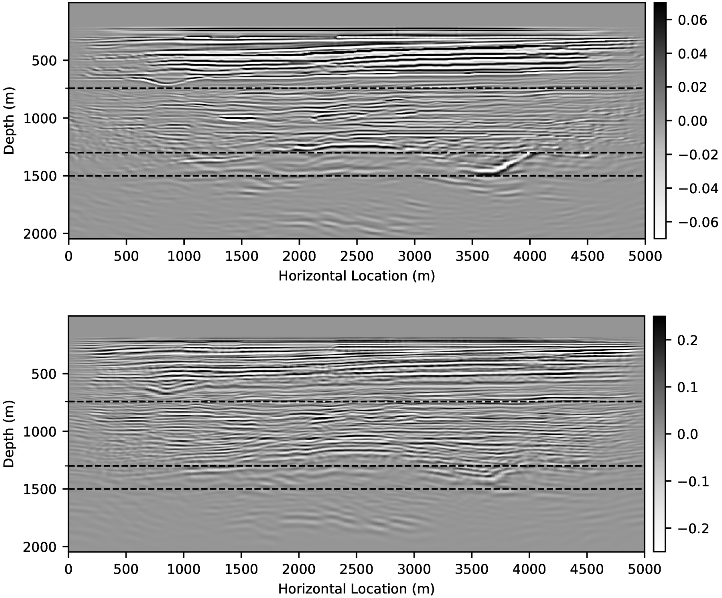
low-fidelity simulation



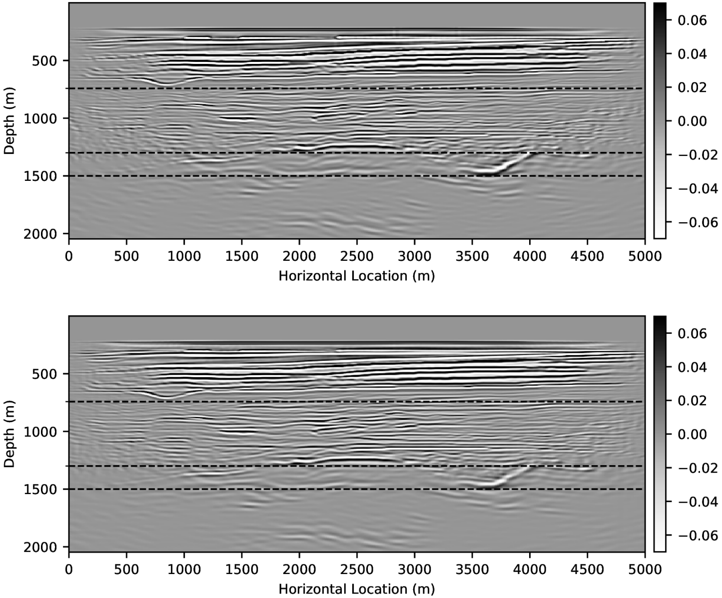
prediction after transfer learning



error w.r.t high-fidelity simulation



(top) high-fidelity image, (bottom) low-fidelity image



(top) high-fidelity image, (bottom) predicted image after transfer learning

Chapter 7. Velocity continuation with Fourier neural operators for accelerated uncertainty quantification

Ali Siahkoohi [1] Mathias Louboutin [1] Felix J. Herrmann [1]

[1]Georgia Institute of Technology

SLIM
Georgia Institute of Technology

Ali Siahkoohi, Mathias Louboutin, and Felix J. Herrmann. "Velocity continuation with Fourier neural operators for accelerated uncertainty quantification". In: *2nd International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. June 2022. URL: https://arxiv.org/abs/2203.14386.

Find $\delta\mathbf{m}$ such that

$$\mathbf{d}_i = \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} + \boldsymbol{\eta}_i$$

linearized Born scattering operator $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$

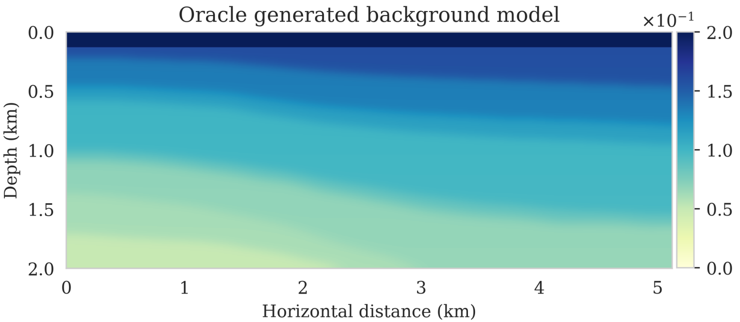
modeling errors $\boldsymbol{\eta}_i$

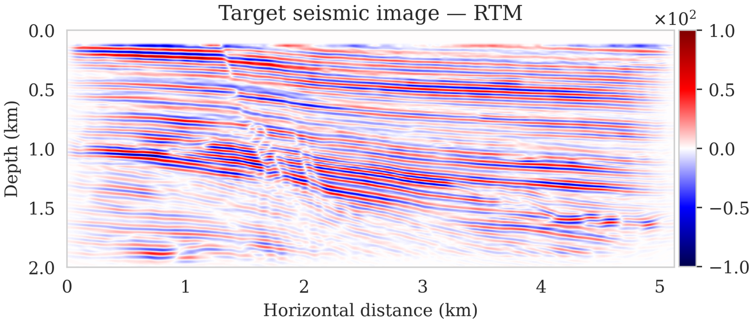
Find $\delta\mathbf{m}$ such that

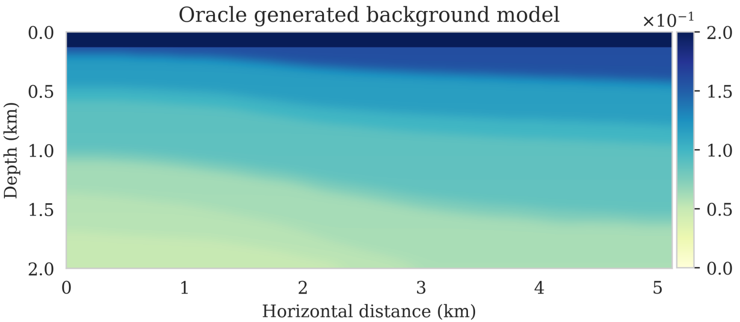
$$\mathbf{d}_i = \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m}$$

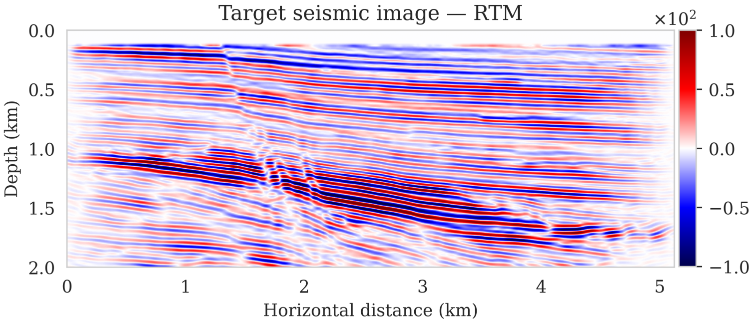
uncertainty (error) in the background model $\mathbf{m}_0$

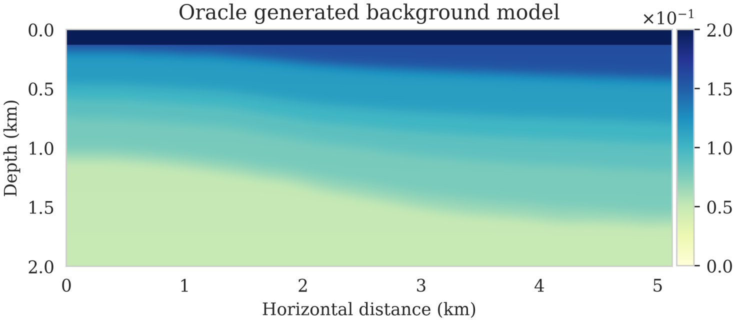
quantifying uncertainty in $\delta\mathbf{m}$ requires solving numerous imaging problems

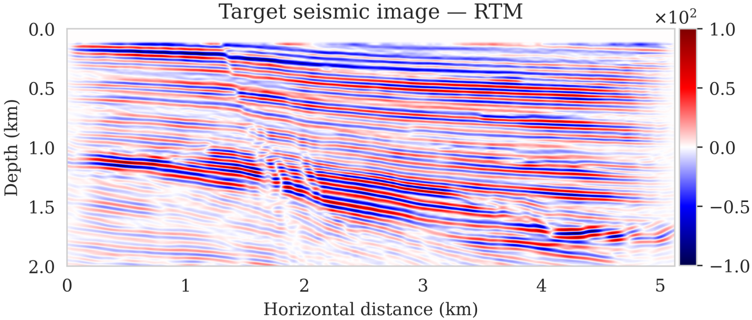

Oracle generated background model
×10⁻¹
0.0
0.5
1.0
1.5
2.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
0.0
0.5
1.0
1.5
2.0


Target seismic image — RTM
×10²
0.0
0.5
1.0
1.5
2.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
1.0
0.5
0.0
−0.5
−1.0


Oracle generated background model
×10⁻¹
Depth (km)
Horizontal distance (km)


Target seismic image — RTM
×10²
Depth (km)
Horizontal distance (km)


Oracle generated background model
×10⁻¹
Depth (km)
Horizontal distance (km)
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5


Target seismic image — RTM
×10²
Depth (km)
Horizontal distance (km)
1.0
0.5
0.0
−0.5
−1.0

Velocity continuation

$$\mathcal{T}_{\left(\mathbf{m}_{\text{init}},\mathbf{m}_{\text{target}}\right)} : \delta\mathcal{M} \rightarrow \delta\mathcal{M}$$

space of seismic images $\delta\mathcal{M}$

initial and target background models $\mathbf{m}_{\text{init}}, \mathbf{m}_{\text{target}}$

maps images associated with one background model to another without explicitly imaging

Sergey Fomel. "Time-migration velocity analysis by velocity continuation". In: *Geophysics* 68.5 (2003), pp. 1662–1672.

Mengmeng Yang, Marie Graff, Rajiv Kumar, and Felix J Herrmann. "Low-rank representation of omnidirectional subsurface extended image volumes". In: *Geophysics* 86.3 (2021), S165–S183.

Contributions

a survey-specific Fourier neural operator surrogate to velocity continuation

mapping images associated with one background model to another virtually for free

trained with only 200 background and seismic image pairs

enables accelerated uncertainty quantification w.r.t background model

Fourier neural operator surrogate

$$\mathcal{G}_{\mathbf{w}^*}(\mathbf{m}_{\text{target}}, \delta\mathbf{m}_{\text{init}}) \approx \mathcal{T}_{(\mathbf{m}_{\text{init}},\mathbf{m}_{\text{target}})}(\delta\mathbf{m}_{\text{init}}) = \delta\mathbf{m}_{\text{target}},$$

Fourier neural operator $\mathcal{G}_{\mathbf{w}}$

initial and target seismic images $\delta\mathbf{m}_{\text{init}}, \delta\mathbf{m}_{\text{target}}$

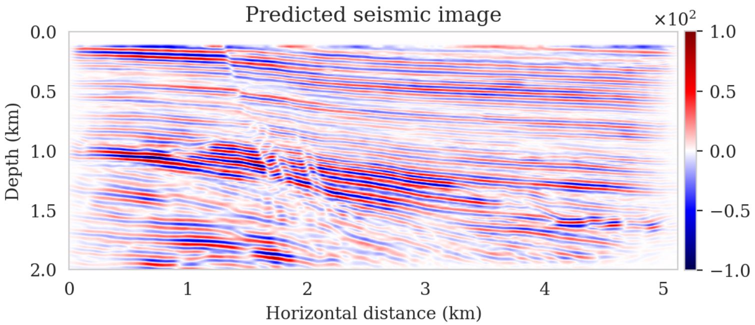
Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. "Fourier Neural Operator for Parametric Partial Differential Equations". In: *9th International Conference on Learning Representations*. OpenReview.net, 2021. URL: https://openreview.net/forum?id=c8P9NQVtmnO.

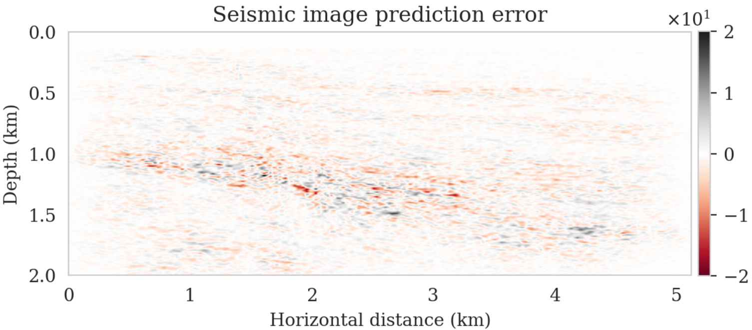
Training

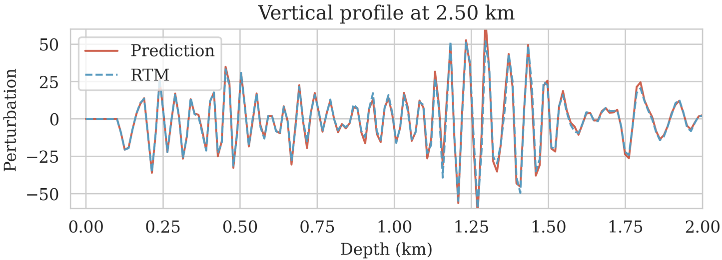
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \frac{1}{n_{\text{train}}} \sum_{i=1} \|\mathcal{G}_{\mathbf{w}}(\mathbf{m}_0^{(i)}, \delta\mathbf{m}_{\text{init}}) - \delta\mathbf{m}_{\text{RTM}}^{(i)}\|_2^2.$$

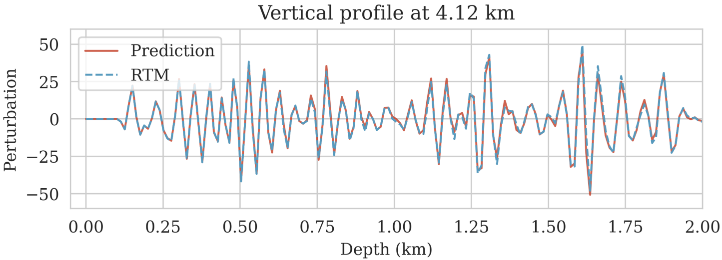
training tuples $\left\{ \left((\mathbf{m}_0^{(i)}, \delta\mathbf{m}_{\text{init}}), \delta\mathbf{m}_{\text{RTM}}^{(i)} \right) \;\middle|\; i = 1, \ldots, n_{\text{train}} \right\}$

$\delta\mathbf{m}_{\text{RTM}} = \sum_{i=1}^{n_s} \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)^\top \mathbf{d}_i.$


Predicted seismic image
×10^2
1.0
0.5
0.0
−0.5
−1.0
Depth (km)
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5
Horizontal distance (km)


Seismic image prediction error
×10^1
2
1
0
−1
−2
Depth (km)
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5
Horizontal distance (km)


Vertical profile at 2.50 km
Prediction
RTM
Perturbation
50
25
0
−25
−50
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Depth (km)


Vertical profile at 4.12 km
Prediction
RTM
Perturbation
50
25
0
−25
−50
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Depth (km)

Main conclusions and contributions

regularizing seismic imaging with deep priors with uncertainty quantification

a systematic approach to translate uncertainty in seismic imaging to downstream tasks

a computationally feasible weak deep prior formulation for large-scale problems

a data-driven variational inference preconditioning scheme

a latent distribution correction method for robust amortized posterior sampling

finite-difference numerical dispersion attenuation with limited in-distribution data

a data-driven surrogate to velocity continuation for accelerated uncertainty quantification

Future directions

deep priors

- MCMC may not be feasible in realistic problems
- variational inference over the weights

reliable amortized variational inference

- theoretically understanding of the limitations of diagonal latent distribution correction
- exploiting amortized variational inference to accelerate MCMC methods
- exploring other probability divergences to address the limitations of KL divergence

data-driven surrogates for PDE solvers

- great progress these days in deep-learning assisted PDE solvers
- finding a right balance b/w existing numerical methods and data-driven methods
- velocity continuation: generalize across a family of survey areas while remaining reliable

My work at SLIM I

Journal papers

Siahkoohi, **Ali**, Mathias Louboutin, and Felix J. Herrmann. "The importance of transfer learning in seismic modeling and imaging". In: *Geophysics* 84.6 (Nov. 2019), A47–A52. DOI: 10.1190/geo2019-0056.1.

Siahkoohi, **Ali**, Gabrio Rizzuti, and Felix J. Herrmann. "Deep Bayesian inference for seismic imaging with tasks". In: *Geophysics* 87.5 (June 2022). DOI: 10.1190/geo2021-0666.1. URL: https://arxiv.org/abs/2110.04825.

Siahkoohi, **Ali**, Gabrio Rizzuti, Rafael Orozco, and Felix J. Herrmann. "Reliable amortized variational inference via physics-based latent distribution correction". In: *To be submitted to Geophysics* (July 2022).

Conference papers

Siahkoohi, **Ali**, Rajiv Kumar, and Felix J. Herrmann. "Seismic Data Reconstruction with Generative Adversarial Networks". In: *80th EAGE Conference and Exhibition*. Extended Abstracts, June 2018. DOI: 10.3997/2214-4609.201801393.

Siahkoohi, **Ali**, Mathias Louboutin, Rajiv Kumar, and Felix J. Herrmann. "Deep-convolutional neural networks in prestack seismic—two exploratory examples". In: *88th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Oct. 2018, pp. 2196–2200. DOI: 10.1190/segam2018-2998599.1.

Herrmann, Felix J., **Ali Siahkoohi**, and Gabrio Rizzuti. "Learned imaging with constraints and uncertainty quantification". In: *Neural Information Processing Systems (NeurIPS) 2019 Deep Inverse Workshop*. Dec. 2019. URL: https://arxiv.org/pdf/1909.06473.pdf.

Rizzuti, Gabrio, **Ali Siahkoohi**, and Felix J. Herrmann. "Learned iterative solvers for the Helmholtz equation". In: *81st EAGE Conference and Exhibition*. Extended Abstracts, June 2019. DOI: 10.3997/2214-4609.201901542.

Siahkoohi, **Ali**, Rajiv Kumar, and Felix J. Herrmann. "Deep-learning based ocean bottom seismic wavefield recovery". In: *89th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Aug. 2019, pp. 2232–2237. DOI: 10.1190/segam2019-3216632.1.

Siahkoohi, **Ali**, Dirk J. Verschuur, and Felix J. Herrmann. "Surface-related multiple elimination with deep learning". In: *89th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Aug. 2019, pp. 4629–4634. DOI: 10.1190/segam2019-3216723.1.

Rizzuti, Gabrio, **Ali Siahkoohi**, Philipp A. Witte, and Felix J. Herrmann. "Parameterizing uncertainty by deep invertible networks, an application to reservoir characterization". In: *90th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Sept. 2020, pp. 1541–1545. DOI: 10.1190/segam2020-3428150.1. URL: https://arxiv.org/abs/2004.07871.

Siahkoohi, **Ali**, Gabrio Rizzuti, and Felix J. Herrmann. "A deep-learning based Bayesian approach to seismic imaging and uncertainty quantification". In: *82nd EAGE Conference and Exhibition*. Extended Abstracts, 2020. DOI: 10.3997/2214-4609.202010770.

My work at SLIM II

Siahkoohi, **Ali**, Gabrio Rizzuti, and Felix J. Herrmann. "Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach". In: *90th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Sept. 2020, pp. 1636–1640. DOI: 10.1190/segam2020-3417560.1. URL: https://arxiv.org/abs/2004.00227.

Siahkoohi, **Ali**, Gabrio Rizzuti, and Felix J. Herrmann. "Weak deep priors for seismic imaging". In: *90th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Sept. 2020, pp. 2998–3002. DOI: 10.1190/segam2020-3417568.1. URL: https://arxiv.org/abs/2004.06835.

Zhang, Mi, **Ali Siahkoohi**, and Felix J. Herrmann. "Transfer learning in large-scale ocean bottom seismic wavefield reconstruction". In: *90th Annual International Meeting*. Society of Exploration Geophysicists. Expanded Abstracts, Sept. 2020, pp. 1666–1670. DOI: 10.1190/segam2020-3427882.1. URL: https://arxiv.org/abs/2004.07388.

Kumar, Rajiv, Maria Kotsi, **Ali Siahkoohi**, and Alison Malcolm. "Enabling uncertainty quantification for seismic data preprocessing using normalizing flows (NF)—An interpolation example". In: *First International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. Expanded Abstracts, 2021, pp. 1515–1519. DOI: 10.1190/segam2021-3583705.1.

Orozco, Rafael, **Ali Siahkoohi**, Gabrio Rizzuti, Tristan van Leeuwen, and Felix J. Herrmann. "Photoacoustic imaging with conditional priors from normalizing flows". In: *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*. Dec. 2021. URL: https://openreview.net/forum?id=woi1OTvROO1.

Ren, Yuxiao, Philipp A. Witte, **Ali Siahkoohi**, Mathias Louboutin, and Felix J. Herrmann. "Seismic Velocity Inversion and Uncertainty Quantification Using Conditional Normalizing Flows". In: *American Geophysical Union (AGU) Fall Meeting*. Dec. 2021. URL: https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/815883.

Siahkoohi, **Ali** and Felix J. Herrmann. "Learning by example: Fast reliability-aware seismic imaging with normalizing flows". In: *First International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. Expanded Abstracts, 2021, pp. 1580–1585. DOI: 10.1190/segam2021-3581836.1. URL: https://arxiv.org/abs/2104.06255.

Siahkoohi, **Ali**, 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.

Louboutin, Mathias, Philipp Witte, **Ali Siahkoohi**, Gabrio Rizzuti, Ziyi Yin, Rafael Orozco, and Felix J. Herrmann. "Accelerating innovation with software abstractions for scalable computational geophysics". In: *2nd International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. June 2022. URL: https://arxiv.org/abs/2203.15038.

My work at SLIM III

Siahkoohi, **Ali**, Michael Chinen, Tom Denton, W. Bastiaan Kleijn, and Jan Skoglun. "Ultra-Low-Bitrate Speech Coding with Pretrained Transformers". In: *Interspeech*. International Speech Communication Association (ISCA). June 2022. URL: https://arxiv.org/abs/2207.02262.

Siahkoohi, **Ali**, Mathias Louboutin, and Felix J. Herrmann. "Velocity continuation with Fourier neural operators for accelerated uncertainty quantification". In: *2nd International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. June 2022. URL: https://arxiv.org/abs/2203.14386.

Siahkoohi, **Ali**, Rafael Orozco, Gabrio Rizzuti, and Felix J. Herrmann. "Wave-equation based inversion with amortized variational Bayesian inference". In: *EAGE Deep learning for seismic processing: Investigating the foundations workshop*. June 2022. URL: https://arxiv.org/abs/2203.15881.

Yin, Ziyi, **Ali Siahkoohi**, Mathias Louboutin, and Felix J. Herrmann. "Learned coupled inversion for carbon sequestration monitoring and forecasting with Fourier neural operators". In: *2nd International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. June 2022. URL: https://arxiv.org/abs/2203.14396.

Zhang, Yijun, Mathias Louboutin, **Ali Siahkoohi**, Ziyi Yin, Rajiv Kumar, and Felix J. Herrmann. "A simulation-free seismic survey design by maximizing the spectral gap". In: *2nd International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists. June 2022. URL: https://arxiv.org/abs/2204.02801.

Technical reports

Siahkoohi, **Ali**, Mathias Louboutin, and Felix J. Herrmann. *Neural network augmented wave-equation simulation*. Tech. rep. Georgia Institute of Technology, Sept. 2019. URL: https://arxiv.org/abs/1910.00925.

Siahkoohi, **Ali**, Gabrio Rizzuti, Philipp A. Witte, and Felix J. Herrmann. *Faster Uncertainty Quantification for Inverse Problems with Conditional Normalizing Flows*. Tech. rep. Georgia Institute of Technology, July 2020. URL: https://arxiv.org/abs/2007.07985.

Louboutin, Mathias, **Ali Siahkoohi**, Rongrong Wang, and Felix J. Herrmann. *Low-memory stochastic backpropagation with multi-channel randomized trace estimation*. Tech. rep. Georgia Institute of Technology, June 2021. URL: https://arxiv.org/abs/2106.06998.

Many thanks to

Felix Herrmann

PhD committee members

colleagues and mentors at SLIM

Chrome Media team at Google

my human and feline families

funding agencies and industry partners