A deep-learning based Bayesian approach to seismic imaging and uncertainty quantification

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann

School of Computational Science and Engineering
Georgia Institute of Technology

February 2020
Inverse problems

Estimate unknown parameters of a system via indirect measurements

- seismology: estimate the speed of sound in subsurface of the Earth
- medical imaging: infer visual representations of the interior of a body (X-ray radiography, MRI)
Inverse problems

Figure 1: A generic inverse problem
Seismic data acquisition

Figure 2: A schematic representation of a seismic survey.\(^1\)

---

Figure 3: Shot record— a collection of traces associated with one source.\(^2\)

\(^2\)Fang, “Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion”. 
Seismic data volume

Figure 4: The entire seismic data obtained from one survey.\textsuperscript{3}

\textsuperscript{3}Fang, “Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion”. 
Objective of exploration seismology

Figure 5: Ultimate goal.\textsuperscript{4}

\textsuperscript{4}Fang, “Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion”. 
Wave speed in the subsurface

Figure 6: Long- and short-wavelength components of velocity model.⁵

⁵Fang, “Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion”. 
Seismic imaging

Figure 7: Goal of seismic imaging.\(^6\)

\(^6\)Fang, “Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion”. 
Nonlinear forward operator

\[ F(m, q_i) = PA(m)^{-1}q_i, \quad i = 1, 2, \ldots, N \]  

where

- \( F(\cdot, \cdot) \): nonlinear forward operator
- \( m \): squared-slowness model
- \( q_i \): source signature of \( i^{th} \) source experiment
- \( P \): restriction operator, restricting wavefields to the receiver locations
- \( A(\cdot) \): discretized wave equation
- \( N \): number of source experiments
Taylor’s series expansion

\[ F(m_0 + \delta m, q_i) = F(m_0, q_i) + \nabla F(m_0, q_i) \delta m + O(\|\delta m\|^2) \]

\[ d^{(\text{obs})}_i = d^{(\text{pred})}_i + J(m_0, q_i) \delta m + O(\|\delta m\|^2) \]  \hspace{1cm} (2)

where

- **\( m_0 \):** background (long-wavelength) squared-slowness model
- **\( \delta m = m - m_0 \):** unknown squared-slowness perturbation model
- **\( d^{(\text{obs})}_i = F(m_0 + \delta m, q_i) \):** observed data
- **\( d^{(\text{pred})}_i = F(m_0, q_i) \):** predicted (simulated) data
- **\( J(m_0, q_i) = \nabla F(m_0, q_i) \):** linear forward operator
- **\( O(\|\delta m\|^2) \):** linearization error
\[ \delta d_i = J(m_0, q_i) \delta m + O(\|\delta m\|^2) \]

unknown linearization error

where

\[ \delta d_i = d_i^{(\text{obs})} - d_i^{(\text{pred})} : \text{data residual} \]
Seismic imaging— challenges

Involves an inconsistent, mildly ill-conditioned linear inverse problem due to:

▶ presence of shadow zones and complex structures in the subsurface
▶ coherent linearization error—i.e., \( \delta d_i = J(m_0, q_i) \delta m + O(\| \delta m \|_2^2) \)
▶ noise in observed data—i.e., \( d_i^{(obs)} = F(m, q_i) + \epsilon_i, \quad \epsilon_i \sim p_{\text{noise}}(\epsilon) \)

Requires prior/regularization. Due to Earth’s heterogeneity:

▶ not possible to precisely encode our prior knowledge
▶ do not have access to samples from ground-truth prior to utilize data-driven priors
Seismic imaging—challenges

Computational challenges

- applying \( J(m_0, q_i) \) or \( J(m_0, q_i)^T \) involves two expensive PDE solves:

\[
J(m_0, q_i) = -PA(m_0)^{-1}[\nabla A(m_0)(A(m_0)^{-1}q_i)]
\]  \hspace{1cm} (4)

- many source experiments—i.e., \( N \) is large
Bayesian inversion

Why Bayesian?
► uncertainty quantification
► incorporating uncertainty into the inversion—e.g., conditional mean estimate

Ultimate goal:
► sampling the posterior distribution

Challenges:
► need for a prior distribution
► expensive to sample the posterior
A Bayesian approach to seismic imaging

Bayes’ rule:

\[ p_{\text{post}} \left( \delta \mathbf{m} \mid \{\delta \mathbf{d}_i, q_i\}_{i=1}^N \right) \propto p_{\text{noise}} \left( \{\delta \mathbf{d}_i, q_i\}_{i=1}^N \mid \delta \mathbf{m} \right) p_{\text{prior}} (\delta \mathbf{m}) \]  

(5)

where

\begin{itemize}
  \item \( p_{\text{post}} \): posterior distribution density
  \item \( p_{\text{noise}} \): density of the noise distribution
  \item \( p_{\text{prior}} \): prior distribution density
\end{itemize}
Prior distribution

Conventional methods—handcrafted and unrealistic priors \(^7\),

- Gaussian or Laplace distribution prior in the physical/transform domain
- tend to bias the outcome of inversion

*Pretrained* generative models as an implicit prior \(^8,9\),

- i.e., requires samples from the *ground-truth* prior distribution
- allows for MCMC sampling in the low-dimensional latent space

---


Prior distribution

*Joint inversion and training* a generative model\(^\text{10}\),

- does not require a pretrained generative model
- fast posterior sampling—feed-forward evaluation of the generative model

**Proposed approach**—an implicit structured *deep prior*\(^\text{11,12,13}\),

- i.e., reparameterize \(\delta m\) w/ a randomly initialized deep CNN
- promotes *natural* images, but not unnatural noise

---


Prior distribution— implicit deep prior

\[ \delta \mathbf{m} = g(\mathbf{z}, \mathbf{w}), \quad \mathbf{w} \sim p_{\text{prior}}(\mathbf{w}) := N(\mathbf{w} \mid 0, \frac{1}{\lambda^2} \mathbf{I}) \tag{6} \]

where

\( g(\cdot, \cdot) \): a randomly initialized deep CNN
\( \mathbf{z} \sim N(\mathbf{0}, \mathbf{I}) \): fixed input to the CNN
\( \mathbf{w} \): unknown CNN weights—e.g., convolutional kernels and biases
\( p_{\text{prior}}(\mathbf{w}) \): Gaussian prior on \( \mathbf{w} \)
\( \lambda \): a hyperparameter
First and second order statistics of the implicit deep prior
Negative log-likelihood

\[- \log p_{\text{noise}} \left( \{ \delta d_i, q_i \}_{i=1}^N \mid w \right) = - \sum_{i=1}^N \log p_{\text{noise}} \left( \delta d_i, q_i \mid w \right) \]

\[= \frac{1}{2\sigma^2} \sum_{i=1}^N \| \delta d_i - J(m_0, q_i)g(z, w) \|_2^2 + \text{const}, \]

independent of \(w\)

where

- \(p_{\text{noise}}\): Gaussian distribution on the noise
- \(\sigma^2\): estimated noise variance
Negative log-posterior

\[- \log p_{\text{post}} \left( \mathbf{w} \mid \{\delta_i, q_i\}_{i=1}^N \right) \]

\[= \frac{1}{2\sigma^2} \sum_{i=1}^{N} \| \delta_i - J(m_0, q_i)g(z, \mathbf{w}) \|^2_2 + \frac{\lambda^2}{2} \| \mathbf{w} \|^2_2 + \text{const, independent of } \mathbf{w} \]  \hspace{1cm} (8)

where

\[p_{\text{post}}: \text{ posterior distribution density on } \mathbf{w}\]
Point estimators

Maximum likelihood estimator:

\[
\delta \hat{m}_{\text{MLE}} = \arg\min_{\delta \hat{m}} \log p_{\text{noise}} \left( \{\delta d_i, q_i\}_{i=1}^{N} \mid \delta \hat{m} \right) \\
= \arg\min_{\delta \hat{m}} \frac{1}{2\sigma^2} \sum_{i=1}^{N} \|\delta d_i - J(m_0, q_i)\delta \hat{m}\|_2^2
\]

- computationally feasible for large-scale problems
- does not regularize, not suitable for ill-posed problems
- does not incorporate uncertainty

---

14 Ozan Oktem. *Bayesian inversion for tomography through machine learning.*
Point estimators

Maximum a-posteriori (MAP) estimator:

\[
\hat{\delta m}_{\text{MAP}} = g(z, \hat{w}_{\text{MAP}}),
\]

where,

\[
\hat{w}_{\text{MAP}} = \arg\min_w -\log p_{\text{post}} \left( w \mid \{\delta d_i, q_i\}_{i=1}^N \right)
\]

\[
= \arg\min_{\delta m} \frac{1}{2\sigma^2} \sum_{i=1}^N \|\delta d_i - J(m_0, q_i)g(z, w)\|_2^2 + \frac{\lambda^2}{2} \|w\|_2^2
\]

► incorporates prior information, suitable for most ill-posed problem
► does not incorporate uncertainty

\[15\]

Oktem, *Bayesian inversion for tomography through machine learning.*
Point estimators

Conditional mean estimator:

\[ \hat{\delta m} := \mathbb{E}_{w \sim p_{\text{post}}(w|\{\delta d_i, q_i\}_{i=1}^N)}[g(z, w)] = \int p_{\text{post}}(w|\{\delta d_i, q_i\}_{i=1}^N)g(z, w)dw \] (11)

► incorporates prior information, suitable for most ill-posed problem
► Bayesian learning reduces overfitting
► incorporates uncertainty into inversion
► involves high-dimensional integration

\(^{16}\)Cheng et al., “A Bayesian Perspective on the Deep Image Prior”.
\(^{17}\)Oktem, Bayesian inversion for tomography through machine learning.
Monte Carlo integration

Approximating the integration by a sum with Monte Carlo integration:

1. sample the posterior distribution, \( \hat{w}_j \sim p_{\text{post}}(w \mid \{\delta d_i, q_i\}_{i=1}^N), \ j = 1, \ldots, T \)

2. approximate the expectation by the mean using the samples,

\[
\mathbb{E}_{w \sim p_{\text{post}}(w \mid \{\delta d_i, q_i\}_{i=1}^N)} [g(z, w)] \approx \frac{1}{T} \sum_{j=1}^T g(z, \hat{w}_j) \tag{12}
\]
Sampling the posterior distribution

Bayesian inference in deep CNNs,

- generally intractable due to high-dimensional parameters space
- a popular approach to sample the posterior is stochastic gradient Langevin dynamics (SGLD)\textsuperscript{18}

\[
\mathbf{w}_{k+1} = \mathbf{w}_k - \frac{\epsilon}{2} \nabla_{\mathbf{w}} L^{(j)}(\mathbf{w}_k) + \eta_k, \quad \eta_k \sim \mathcal{N}(0, \epsilon I),
\]  \hspace{1cm} (13)

where

- $\epsilon$: stepsize
- $L^{(j)}(\mathbf{w}) = \frac{N}{2\sigma^2} \| \delta \mathbf{d}_j - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_j) \mathbf{g}(\mathbf{z}, \mathbf{w}) \|^2_2 + \frac{\lambda^2}{2} \| \mathbf{w} \|^2_2$ approximates the negative-log posterior

Uncertainty quantification

Pointwise standard deviation as a measure of uncertainty,

i.e., pointwise standard deviation among $\hat{w}_j \sim p_{\text{post}}(w | \{\delta d_i, q_i\}_{i=1}^N)$

We expect to see more uncertainty in regions that are more difficult to image—e.g.,

- location of the reflectors
- deeper parts of the model
- close to boundaries and fault zone
Numerical experiment—setup

Synthetic dataset simulated by solving the acoustic wave equation,

- 2D Overthrust velocity model
- finite-difference simulations w/ Devito \(^{19,20}\)
- 369 shot records w/ 369 receivers
- 27 m source/receiver sampling
- 2 seconds recording time
- Ricker source wavelet w/ 8 Hz central frequency


Numerical experiment—setup

Simultaneous source experiments,\(^\text{21}\),

- an economic way to sample seismic data in practice
- wave equation is \textit{linear} in source—i.e., \( F(m, q) = PA(m)^{-1}q \),
- linearly combine observed data w/ normally distributed weights
- generate associated forward operators, \( J(m_0, q) \), with \textit{random superposition} of source signatures

Numerical experiment— setup

Noise in the data

- inherent (non-Gaussian) linearization error— directly estimated variance = 0.490
- measurement noise drawn from $\mathcal{N}(0, 2I)$
- overall signal-to-noise ratio of data = $-11.37$ dB

Deep learning framework

- PyTorch library
- integrate Devito’s linear forward operator into PyTorch— allowing us to compute the gradients w/ automatic differentiation.

---

Numerical experiment—squared-slowness model
Measurement-noise-free data residual

\[ \delta d_{92} \]

\[ \delta d_{184} \]

\[ \delta d_{276} \]
Data residual w/ measurement noise
Measurement-noise-free simultaneous source data residual
Simultaneous source data residual w/ measurement noise
Seismic imaging—MLE

True perturbation model - $\delta m$

MLE (no prior)
Seismic imaging—MAP estimate

True perturbation model - $\delta m$

MAP estimate, $g(z, \hat{w}_{MAP})$
Seismic imaging—conditional mean estimate
Samples from the posterior distribution
Samples from the posterior distribution
Samples from the posterior distribution

\[ g(z, \hat{w}_{83}) \]

\[ g(z, \hat{w}_{83}) - \delta \hat{m} \]
Samples from the posterior distribution

\[ g(z, \hat{\omega}_{89}) \]

\[ g(z, \hat{\omega}_{89}) - \delta \hat{m} \]
Samples from the posterior distribution

\[ g(z, \hat{w}_{90}) \]

\[ g(z, \hat{w}_{90}) - \delta \hat{m} \]
Samples from the posterior distribution

\[ g(z, \hat{w}_{114}) \]

- Depth (km)
- Horizontal distance (km)

\[ g(z, \hat{w}_{114}) - \delta\hat{m} \]

- Depth (km)
- Horizontal distance (km)
Samples from the posterior distribution
Samples from the posterior distribution
Samples from the posterior distribution

\[ \mathbf{g}(\mathbf{z}, \hat{\mathbf{w}}_{128}) \]

\[ \mathbf{g}(\mathbf{z}, \hat{\mathbf{w}}_{128}) - \delta \hat{\mathbf{m}} \]
Samples from the posterior distribution
Samples from the posterior distribution

Point-wise standard deviation of $g(z, \hat{w}_j)$'s

Point-wise standard deviation vertical profiles
Imaging and uncertainty quantification

Point-wise histogram at (1.60 km, 1.98 km)

Point-wise histogram at (8.35 km, 1.62 km)
Observations and conclusions

- Utilized a structured prior induced by a carefully designed CNN
- Capable of sampling the posterior by running SGLD, albeit being expensive
- Jointly captured uncertainty in the imaging and the reparametrization with a CNN
Observations and conclusions

- Deep prior was partially able to circumvent the imaging artifacts
- The conditional mean demonstrated less artifacts
- Pointwise standard variation coincided with regions that are more difficult to image
Paper and code


Code: https://github.com/alisiahkoohi/seismic-imaging-with-SGLD