Training data for Geophysical Inversion

Jodi Mead, Department of Mathematics, *Boise State University*
Koffivi Fidele Gbagbe, *African Institute of Mathematical Sciences*
Amrina Ferdous, Computing PhD (Data Science), *Boise State University*

BOISE STATE UNIVERSITY

Forward Problem

$$\mathbf{d} = \mathbf{G}(\mathbf{m}) + \boldsymbol{\epsilon}$$

- $\mathbf{G} : \mathbb{R}^N \to \mathbb{R}^n$ a mathematical model that may be an analytical or differential equation, an algorithm or a "black box" with inputs and outputs.
- $\mathbf{m} \in R^N$ - parameters in the system
- $\mathbf{d} \in R^n$ - results, or observations of the system
- $\boldsymbol{\epsilon} \in R^n$ - error from obtaining observations/results or failure of the model to describe the system

Inverse Problem

Parameter estimation:

$$\mathbf{m} = \mathbf{G}^{-1}(\mathbf{d} + \boldsymbol{\epsilon})$$

- $\mathbf{G}^{-1} : \mathbb{R}^n \to \mathbb{R}^N$ represents the inverse operator

However in practice:

$$\hat{\mathbf{m}} = \arg\min \|\mathbf{d} - \mathbf{G}(\mathbf{m})\|$$

This is typically an ill-posed problem

- different causes sometimes lead to the same effects
- small change in the data can result in a significant change in the estimated parameters

Bayesian Approach

Given the *prior* distribution $\rho(\mathbf{m})$ and conditional distribution on the data $f(\mathbf{d}|\mathbf{m})$, find the *posterior* distribution $q(\mathbf{m}|\mathbf{d})$:

$$q(\mathbf{m}|\mathbf{d}) = \frac{f(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})}{f(\mathbf{d})}$$

or

$$q(\mathbf{m}|\mathbf{d}) \propto f(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})$$

The mean of $q(\mathbf{m}|\mathbf{d})$ is the MAP estimate and the variance $q(\mathbf{m}|\mathbf{d})$ gives uncertainty in the estimates.

Multivariate Gaussian Distribution

Given

- *prior* distribution $\mathbf{m} \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_m)$
- conditional data distribution $\mathbf{d}|\mathbf{m} \sim \mathcal{N}(\mathbf{G}(\mathbf{m}), \mathbf{C}_d)$

Bayes Rule

$$q(\mathbf{m}|\mathbf{d}) \propto \exp\left\{-1/2(((\mathbf{d}-\mathbf{G}(\mathbf{m}))^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{G}(\mathbf{m})) + (\mathbf{m}-\mathbf{m}_0)^T\mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_0))\right\}$$

MAP Parameter Estimates

Gauss-Newton

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \left(\mathbf{J}_k^T \mathbf{C}_d^{-1} \mathbf{J}_k + \mathbf{C}_m^{-1}\right)^{-1} \left(\mathbf{J}_k^T \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{G}(\mathbf{m}_k)) - \mathbf{C}_m^{-1}(\mathbf{m}_k - \mathbf{m}_0)\right)$$

Assumes:

- Good estimate $\mathbf{m}_0$
- $\mathbf{G}(\mathbf{m})$ is differentiable near $\mathbf{m}_k$
- $\mathbf{J}$ is available
- Hessian $\mathbf{H} \approx \mathbf{J}^T \mathbf{J}$

Posterior Sampling

MCMC

- use Metropolis Hastings algorithm to sample
- create a Markov Chain that doesn't have memory
- reduce bias in $\mathbf{m}_0$ by burning in
- reduce autocorrelation by skipping

Guaranteed to converge to posterior however, for large number of parameters, efficiency is a limiting factor.

Neural Operator Estimate G^{-1}

- Find a surrogate operator $\mathbf{G}^{-1}$ such that $\mathbf{m} = \mathbf{G}^{-1}(\mathbf{d} - \boldsymbol{\epsilon})$
- Universal approximation theorem: a feed-forward network with a single hidden layer having a finite number of artificial neurons can approximate any continuous function
- Recent advances in seismic imaging, electrical impedance tomography, for example

Supervised Learning

- With a good understanding of the probability distribution of $\mathbf{m}$, generate synthetic data using the forward model
- Allows us to get beyond real-world data collecting and usage limitations, and generate as much data as we need

Synthetic Training Data Sets

Two approaches to generating synthetic training data:

1. Bayesian Prior sampling

$$\mathbf{m} \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_m)$$

2. Rejection Posterior sampling

$$q(\mathbf{m}|\mathbf{d}) \propto \exp\left\{-1/2((\mathbf{d} - \mathbf{G}(\mathbf{m}))^T \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{G}(\mathbf{m})))\right\}$$

Rejection Sampling

1. Choose $\mathbf{m}$ from a proposal distribution $g(\mathbf{m})$ e.g.

$$\exp\left\{-1/2(\boldsymbol{\epsilon}^T \mathbf{C}_d^{-1} \boldsymbol{\epsilon})\right\}, \ \overline{\boldsymbol{\epsilon}} = \mathbf{0}, \mathbf{C}_d = \mathbf{I}$$

2. Choose constant c and reject choice $\mathbf{m}$ in 1. if

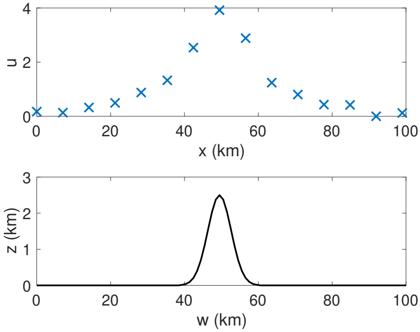
$$\sup_{\mathbf{m}} f(\mathbf{m})/g(\mathbf{m}) < c$$

where $f(\mathbf{m})$ is the target distribution e.g.

$$\exp\left\{-1/2((\mathbf{d} - \mathbf{G}(\mathbf{m}))^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{G}(\mathbf{m})))\right\}$$

Recovering subsurface frontier

$$u(x) = \int_a^b \log \frac{(x-w)^2 + H^2}{(x-w)^2 + (H - z(w))^2} \, dw$$



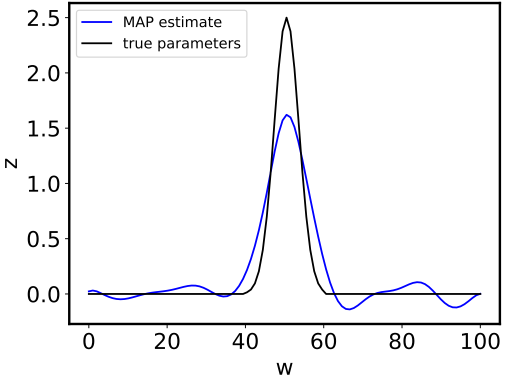
Inverse Problem

$$\mathbf{u} = \mathrm{G}(\mathbf{z}) + \boldsymbol{\epsilon}$$

- $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_d^2\mathbf{I})$, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_u)$
- $C_u(w, w') = \sigma_u^2 \exp\{-1/2(w - w')^2\}$
- $\boldsymbol{\epsilon}, \mathbf{u} \in \mathbb{R}^{15}$
- $\mathbf{z} \in \mathbb{R}^{100}$

MAP Estimate

$\sigma_d = 0.1,\ \sigma_u = 1$

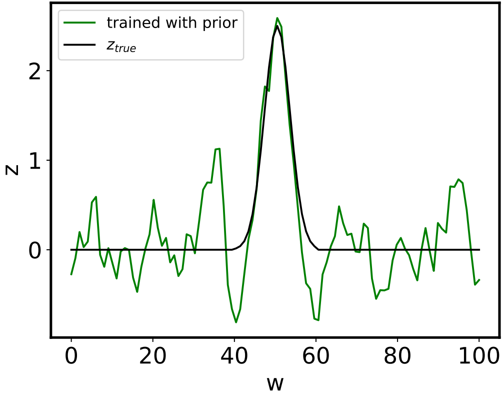


Neural Network

- Multilayer perceptron - 3 layers
- Adam algorithm optimizer
- 1000 synthetic training data, 20% test/train split

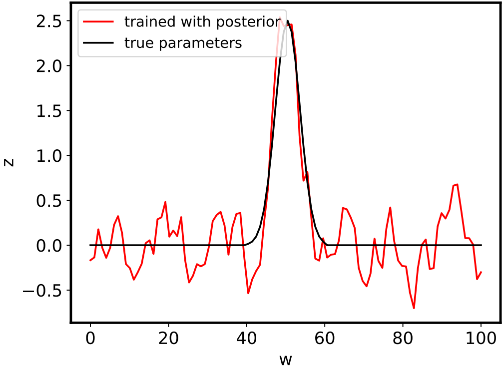
NN Operator Estimate - Prior sampling

$\sigma_d = 0.1$, $\sigma_u = 1$



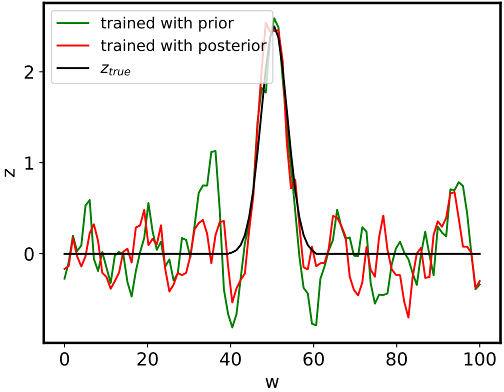
NN Operator Estimate - Posterior sampling

$\sigma_d = 0.1, \; \sigma_u = 1$



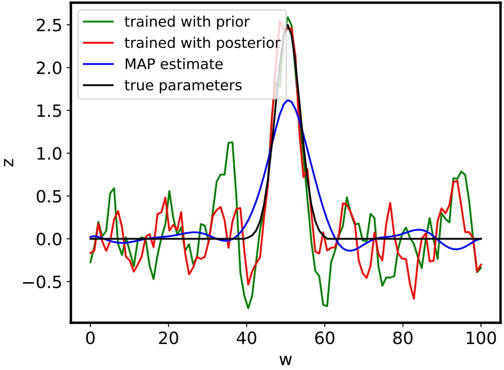
Comparison of NN Operator Estimates

$\sigma_d = 0.1,\ \sigma_u = 1$



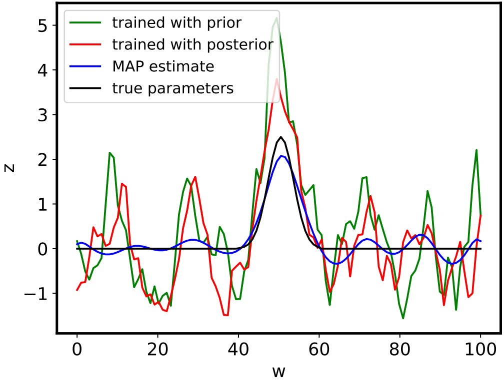
Comparison of MAP and NN Operator Estimates

$\sigma_d = 0.1$, $\sigma_u = 1$



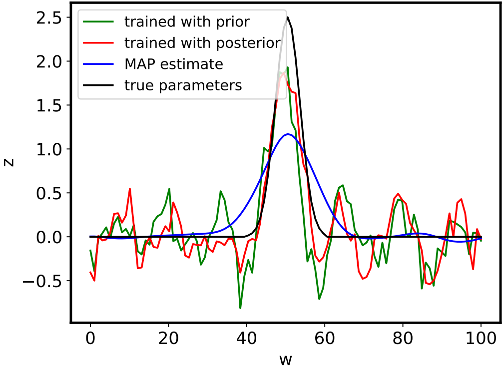
Comparison of MAP and NN Operator Estimates

$\sigma_d = 0.1,\ \sigma_u = 2.5$



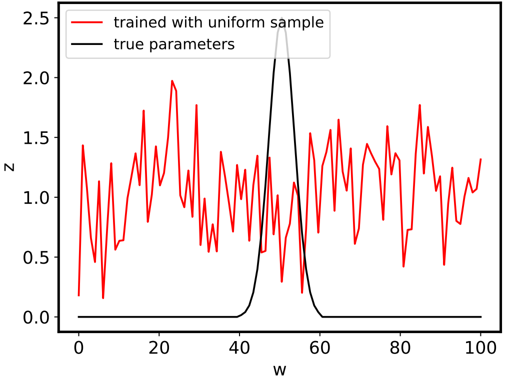
Comparison of MAP and NN Operator Estimates

$\sigma_d = 0.1,\ \sigma_u = 0.5$



Summary and Conclusions

- Successful learning of inverse operators for sufficiently large training data sets has been achieved in seismic imaging, electrical impedance tomography among other applications
- Here we experiment with sampling strategies that create a synthetic data set for a nonlinear ill-posed inverse problem
  - Synthetic training data sampled **uniformly** could not recover the frontier between two densities
  - Sampling the **prior** recovered the frontier, and is noisy
  - **Posterior** sampling, using rejection, slightly improved the recovery found with prior sampling
  - Choice of σ_z in the sampling strategy significantly impacted quality of the results from all methods


trained with uniform sample
true parameters
2.5
2.0
1.5
1.0
0.5
0.0
z
0
20
40
60
80
100
w