Learned imaging with constraints and uncertainty quantification

Felix J. Herrmann, Ali Siahkoohi, and Gabrio Rizzuti. Learned imaging with constraints and uncertainty quantification. In: NeurIPS 2019 Deep Inverse Workshop. 2019

Presented by: Ali Siahkoohi

SLIM
Georgia Institute of Technology

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

unknown parameters

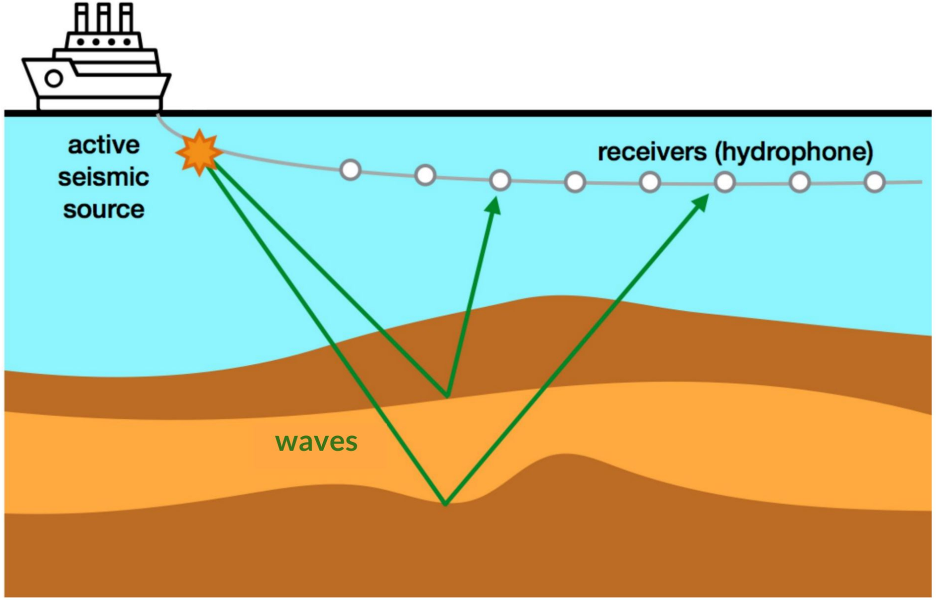
Forward operator

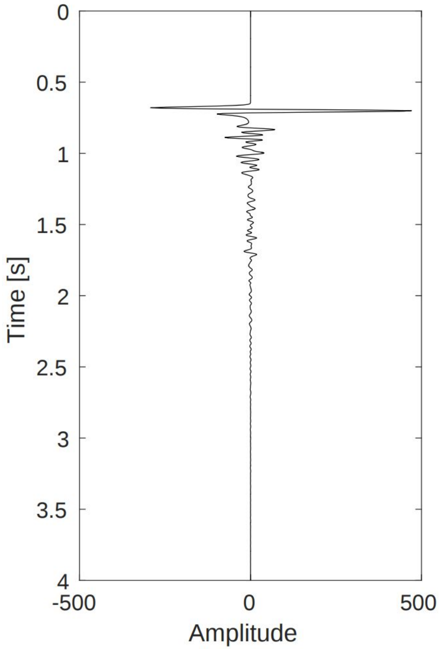
indirect measurements

Inverse problem

Seismic imaging

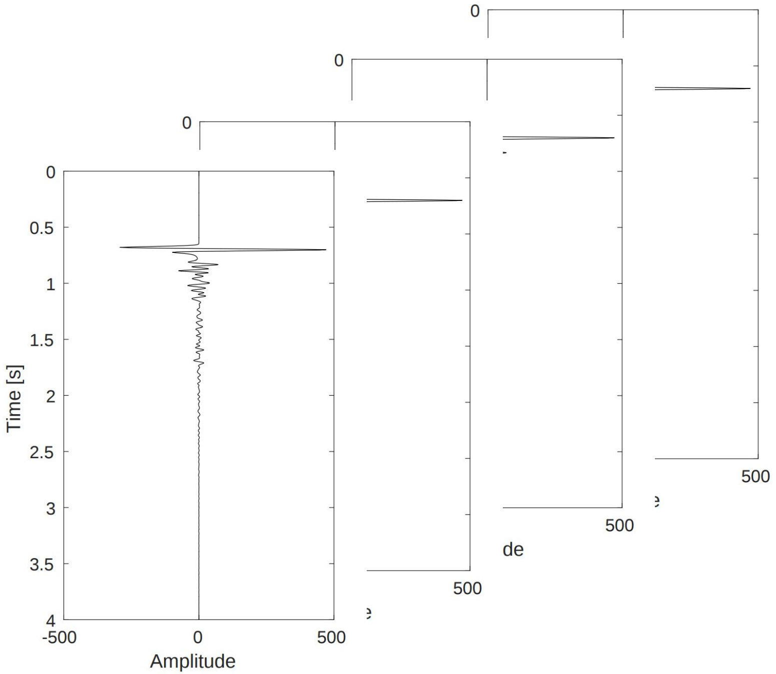
Seismic data acquisition

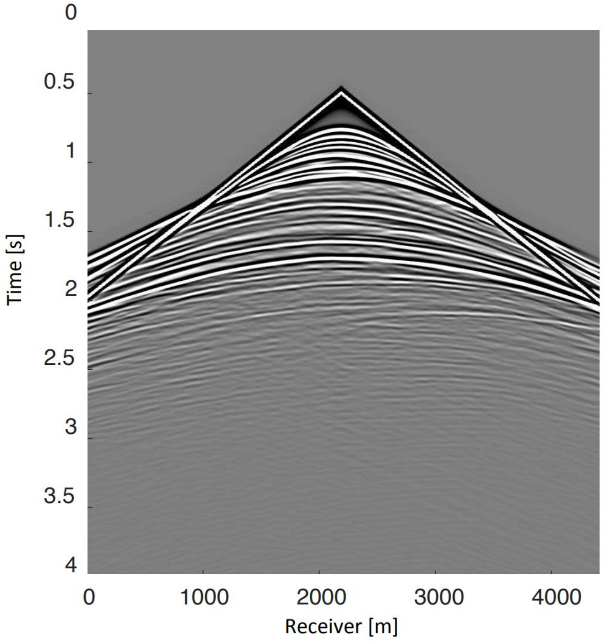




Fang, Z., 2018. Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion (Doctoral dissertation, University of British Columbia).

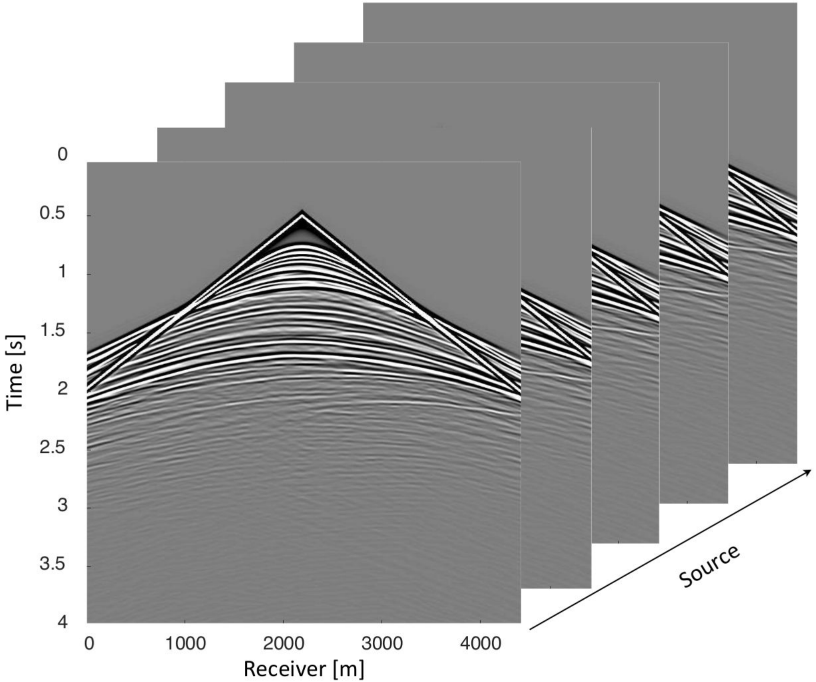
Shot record - one source location





Fang, Z., 2018. Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion (Doctoral dissertation, University of British Columbia).

Seismic data volume



Seismic imaging

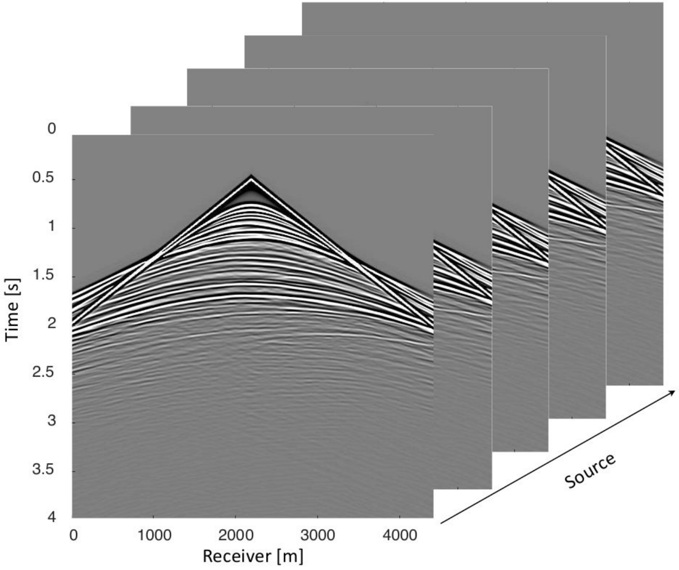
unknown parameters

Forward operator

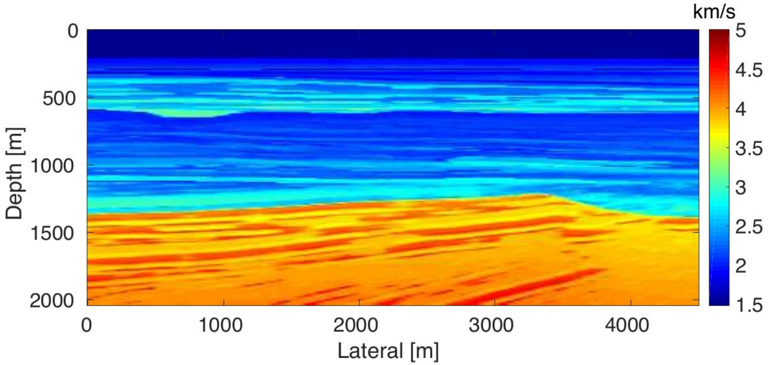
indirect measurements

Seismic imaging

Objective of exploration seismology



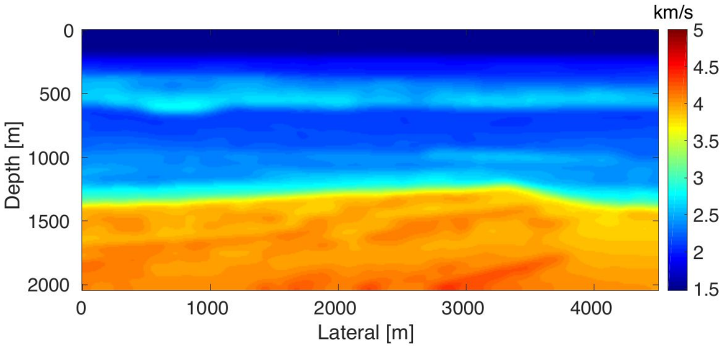
Observed data



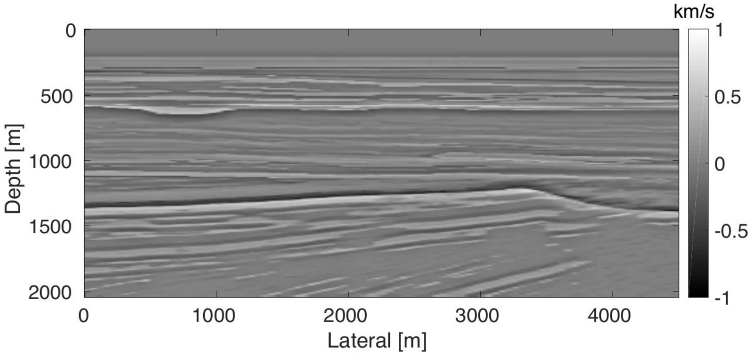
Subsurface velocity structure

Fang, Z., 2018. Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion (Doctoral dissertation, University of British Columbia).

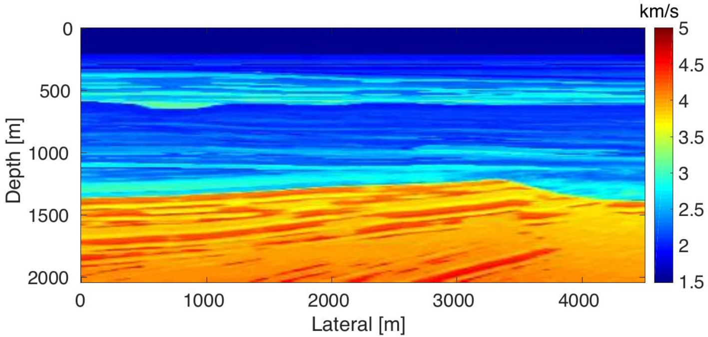
Velocity model



Long-wavelength structure (background velocity)

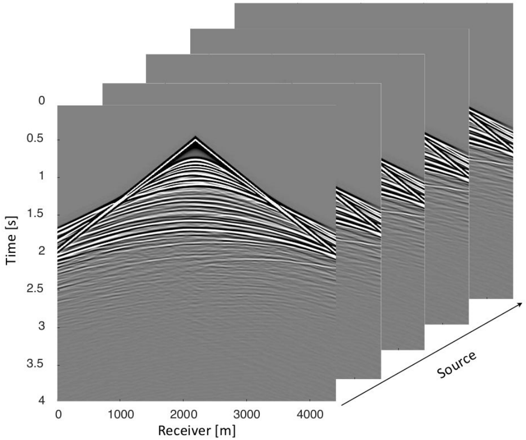


Short-wavelength structure (velocity perturbation)

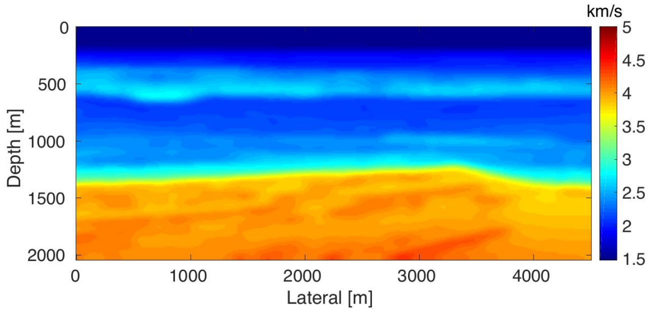


Subsurface velocity structure

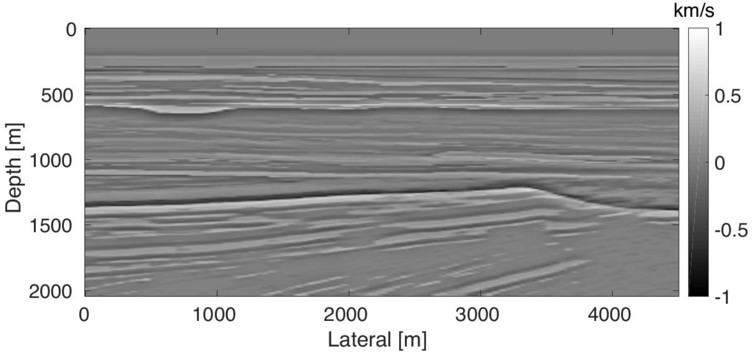
Seismic imaging



Observed data



Long-wavelength structure (background velocity)



Short-wavelength structure (velocity perturbation)

Seismic imaging

unknown parameters

Forward operator

indirect measurements

Seismic imaging

Nonlinear forward operator

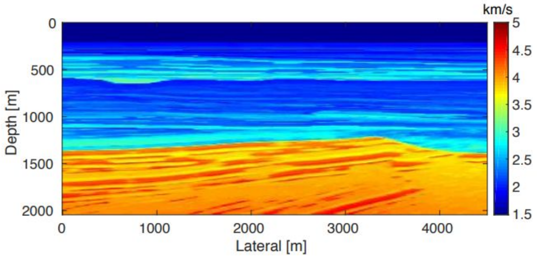
$$F(\mathbf{m},\mathbf{q}) = \mathbf{P}\overbrace{\mathbf{A}(\mathbf{m})}^{\text{discretized wave-equation}}{}^{-1}\mathbf{q}$$

squared slowness

source

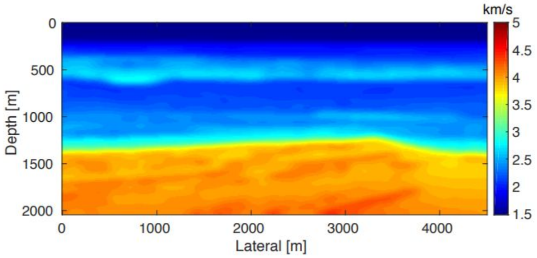
Restriction operator to the receivers location

Taylor series expansion



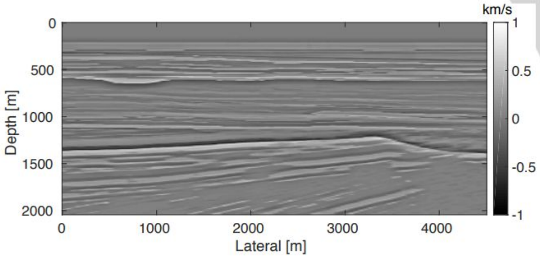
True model

=



Smooth background model

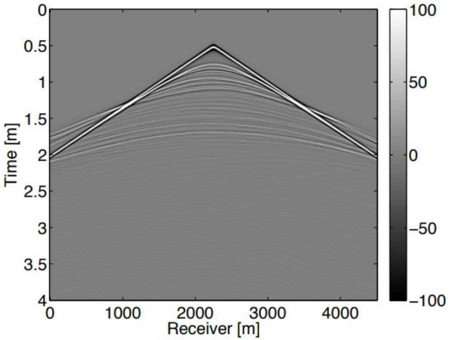
+



Model perturbation

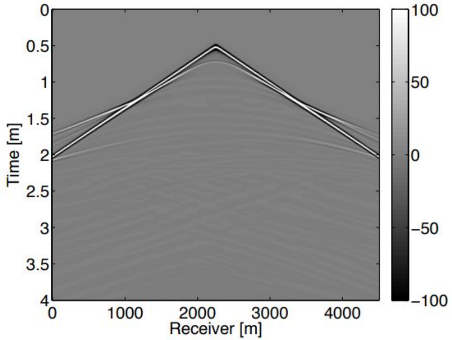
$$F(\mathbf{m}_0 + \delta\mathbf{m}, \mathbf{q}) = F(\mathbf{m}_0, \mathbf{q}) + \nabla F(\mathbf{m}_0, \mathbf{q})^T \delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2)$$

Taylor series expansion



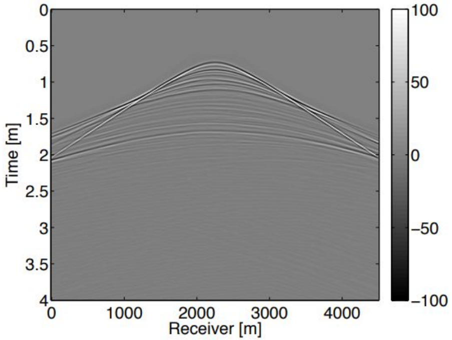
Observed data

=



Predicted data

+

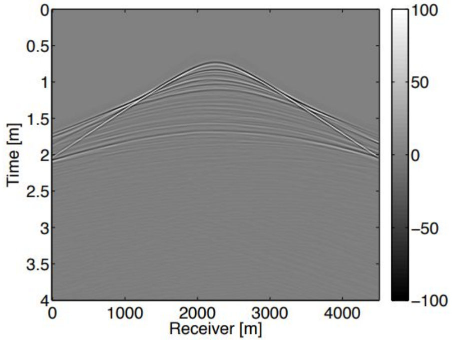


Data residual

$$\overbrace{F(\mathbf{m}_0 + \delta\mathbf{m}, \mathbf{q})} = \overbrace{F(\mathbf{m}_0, \mathbf{q})} + \overbrace{\nabla F(\mathbf{m}_0, \mathbf{q})^T \delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2)}$$

Taylor series expansion

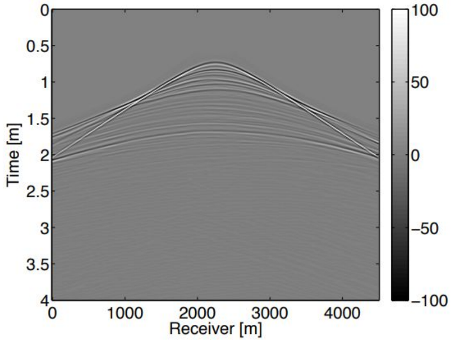
$$\nabla F(\mathbf{m}_0, \mathbf{q})^T \delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2) \quad = $$



Data residual

Linearization error

$$\nabla F(\mathbf{m}_0, \mathbf{q})^T \delta\mathbf{m} \quad \simeq$$



Data residual

Seismic imaging

$$\nabla F(\mathbf{m}_0, \mathbf{q})^T \delta\mathbf{m} \simeq \delta\mathbf{d}$$

Linearized born forward modeling operator

$$\mathbf{J}(\mathbf{m}_0, \mathbf{q}) \equiv \nabla F(\mathbf{m}_0, \mathbf{q})^T$$

Seismic imaging

unknown parameters

Forward operator

indirect measurements

Seismic imaging

Least squares seismic imaging

$$\underset{\delta\mathbf{m}}{\arg\min} \quad \frac{1}{2}\sum_{i=1}^{N} \|\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2 \quad \text{subject to} \quad \delta\mathbf{m} \in \mathcal{C}$$

$$\text{where,} \quad \delta\mathbf{d}_i = \mathbf{d}_{\text{obs},i} - F(\mathbf{m}_0, \mathbf{q}_i)$$

N	Number of source experiments
$\mathbf{q}_i$	Source signature in i^{th} source experiments
$\mathbf{d}_{\text{obs},i}$	Observed data in i^{th} source experiments
$\delta\mathbf{d}_i$	Data residual in i^{th} source experiments
$\mathcal{C}$	A constraint set encoding our prior knowledge

Computational challenges

Applying $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$ and $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)^T$ is expensive,

- Involves two PDE solves:

$$\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i) = \nabla F(\mathbf{m}_0, \mathbf{q})^T = -\mathbf{P}\overbrace{\mathbf{A}(\mathbf{m}_0)^{-1}}\left[\nabla \mathbf{A}(\mathbf{m}_0)\left(\overbrace{\mathbf{A}(\mathbf{m}_0)^{-1}\mathbf{q}}\right)\right]$$

Many source experiments,

- N is large
- Evaluating the objective function requires $2N$ PDE solves.

Need for regularization/prior

Solving inconsistent system of equations,

- linearization error, i.e., $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2) = \delta\mathbf{d}_i$
- noise in observed data, i.e., $\mathbf{d}_{\text{obs},i} = F(\mathbf{m}, \mathbf{q}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim p_{noise}(\boldsymbol{\epsilon})$

Good choices for regularization/prior is challenging,

- Hand crafted priors bias solution towards handcrafted priors

Witte, P.A., Louboutin, M., Luporini, F., Gorman, G.J. and Herrmann, F.J., 2019. Compressive least-squares migration with on-the-fly Fourier transforms. Geophysics, 84(5), pp.1-76.

SLIM

Stochastic optimization edit

Stochastic optimization (over source experiments)

- Approximate objective/gradient with a minibatch of source experiments

$$\sum_{i=1}^{N} \|\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2 \simeq \frac{N}{n}\sum_{i=1}^{n} \|\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2$$

- Typically we need 2-3 passes over the entire set of source experiments

Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A.Y. and Herrmann, F.J., 2018. Total variation regularization strategies in full-waveform inversion. SIAM Journal on Imaging Sciences, 11(1), pp.376-406.
Peters, B., Smithyman, B.R. and Herrmann, F.J., 2019. Projection methods and applications for seismic nonlinear inverse problems with multiple constraints. Geophysics, 84(2), pp.R251-R269.

SLIM

Regularization/prior

Handcrafted quadratic penalty terms

- Tikhonov regularization, a.k.a (weighted) ridge regression

$$\underset{\delta\mathbf{m}}{\arg\min} \quad \frac{1}{2}\sum_{i=1}^{N} \|\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{R}\delta\mathbf{m}\|_2^2$$

- may adversely affect gradient & Hessian
- no guarantees that all model iterates are (physically) feasible
- biases solution towards handcrafted prior

Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A.Y. and Herrmann, F.J., 2018. Total variation regularization strategies in full-waveform inversion. SIAM Journal on Imaging Sciences, 11(1), pp.376-406.
Peters, B., Smithyman, B.R. and Herrmann, F.J., 2019. Projection methods and applications for seismic nonlinear inverse problems with multiple constraints. Geophysics, 84(2), pp.R251-R269.

SLIM

Regularization/prior

Handcrafted constraints

- Total variation constraint

$$\underset{\delta\mathbf{m}}{\arg\min} \quad \frac{1}{2}\sum_{i=1}^{N}\|\mathbf{J}(\mathbf{m}_0,\mathbf{q}_i)\delta\mathbf{m}-\delta\mathbf{d}_i\|_2^2 \quad \text{subject to} \quad \delta\mathbf{m}\in\mathcal{C}$$

- (physical) constraints are satisfied during every iteration
- biases solution towards handcrafted prior

SLIM

Generative models as priors

If we have access to sufficient samples from the ground truth distribution

- train a generative model, e.g. GAN, VAE, to sample the distribution

$$\boldsymbol{x} \sim p_X^{\text{gen}}(\boldsymbol{x}) \Rightarrow \boldsymbol{x} = \boldsymbol{g}(\boldsymbol{z}), \boldsymbol{z} \sim p_Z(\boldsymbol{z})$$

$$p_X^{\text{gen}}(\boldsymbol{x}) = p_X^{\text{true}}(\boldsymbol{x})$$

- use the *pre-trained* generative model as prior

Bayesian Inference

Statistical approach for formulating & solving inverse problems

- quantifies uncertainty in inference and motivates regularization

$$
\begin{aligned}
p_X^{\text{post}}(\boldsymbol{x}|\boldsymbol{y}) &= \frac{1}{\mathbb{Z}} p^{\text{l}}(\boldsymbol{y}|\boldsymbol{x}) p_X^{\text{prior}}(\boldsymbol{x}) \\
&= \frac{1}{\mathbb{Z}} p_\eta(\hat{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x})) p_X^{\text{prior}}(\boldsymbol{x})
\end{aligned}
$$

p_X^{prior}: prior distribution.

p_η: distribution for the noise in measurement

Inference with generative model

Use the *pre-trained* generative model as prior in Bayesian inference

$$p_X^{\text{post}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{\mathbb{Z}} p_\eta(\hat{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x})) p_X^{\text{gen}}(\boldsymbol{x})$$

Performing inference for an arbitrary function:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_X^{\text{post}}}[l(\boldsymbol{x})] &= \frac{1}{\mathbb{Z}} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_X^{\text{gen}}}[l(\boldsymbol{x}) p_\eta(\hat{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x}))], \\
&= \frac{1}{\mathbb{Z}} \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_Z}[l(\boldsymbol{g}(\boldsymbol{z})) p_\eta(\hat{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{z})))], \\
&= \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_Z^{\text{post}}}[l(\boldsymbol{g}(\boldsymbol{z}))],
\end{aligned}$$

Approximate inference

Drawing samples from the posterior can be done by performing MCMC on (more on this later)

$$p_Z^{\text{post}}(\boldsymbol{z}|\boldsymbol{y}) \equiv \frac{1}{\mathbb{Z}} p_\eta(\hat{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{z}))) p_Z(\boldsymbol{z})$$

i.e., $p_Z^{\text{mcmc}}(\boldsymbol{z}|\boldsymbol{y}) \approx p_Z^{\text{post}}(\boldsymbol{z}|\boldsymbol{y})$. Next, for any function:

$$\overline{l(\boldsymbol{x})} \equiv \mathop{\mathbb{E}}_{\boldsymbol{x} \sim p_X^{\text{post}}} [l(\boldsymbol{x})] \approx \frac{1}{N_{\text{samp}}} \sum_{n=1}^{N_{\text{samp}}} l(\boldsymbol{g}(\boldsymbol{z})), \qquad \boldsymbol{z} \sim p_Z^{\text{mcmc}}(\boldsymbol{z}|\boldsymbol{y}).$$

Why generative model as prior?

The prior is no longer hand-crafted

- solution of the inverse problem is not biased towards the prior
  - Only if $p_X^{\text{gen}}(\boldsymbol{x}) = p_X^{\text{true}}(\boldsymbol{x})$
- performing MCMC on the latent variable is easier
  - much smaller dimension

Big assumption: we have access to samples from the ground truth distribution

- limits the application in seismic inversion/imaging

Limitations in seismic inverse problems

Due to Earth's heterogeneity, we do not have access to samples from true prior

$$\{\boldsymbol{x}_i\}_{i=1}^{N} \sim p_{prior}(\boldsymbol{x})$$

i.e., limited chance of pre-training a generative model

We have many source experiments $\mathbf{d}_{\mathrm{obs},i}$ explaining one and only one unknown,

i.e., $(\boldsymbol{y}_1, \boldsymbol{x}), \ (\boldsymbol{y}_2, \boldsymbol{x}), \ \ldots, \ (\boldsymbol{y}_N, \boldsymbol{x})$

Ulyanov, D., Vedaldi, A. and Lempitsky, V., 2018. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 9446-9454).

Deep Image Prior

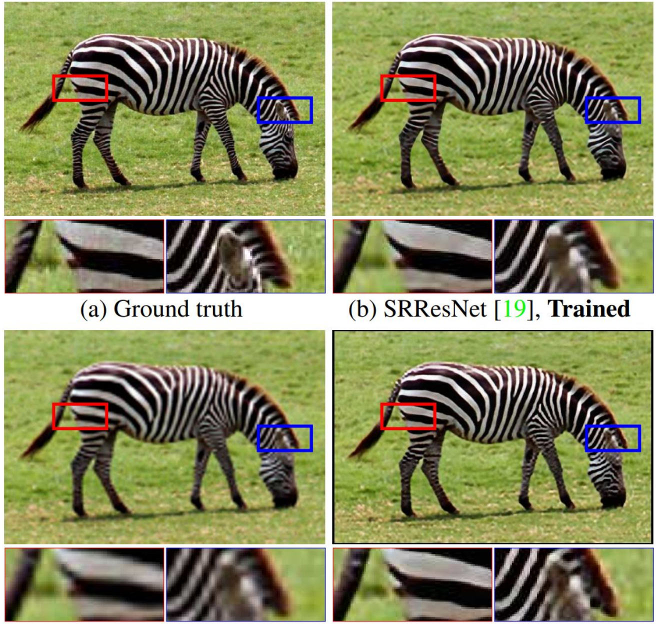
An *untrained* neural net as a prior

$$w^* = \arg\min_w \|y - AG(z;w)\|^2$$

where $y = Ax^* + \eta$ is some noisy observation

$z \in \mathbb{R}^k$ fixed random vector

$G(z;w)\colon \mathbb{R}^k \to \mathbb{R}^n$ *Untrained network*



(a) Ground truth (b) SRResNet [19], **Trained**

(c) Bicubic, **Not trained** (d) Deep prior, **Not trained**

Ulyanov, D., Vedaldi, A. and Lempitsky, V., 2018. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 9446-9454).
Herrmann, F.J., Siahkoohi, A. and Rizzuti, G., 2019. Learned imaging with constraints and uncertainty quantification. arXiv preprint arXiv:1909.06473.

SLIM

Deep Image Prior

$$\underset{w}{\text{minimize}} \ \frac{1}{2}\|y - Ag(z,w)\|_2^2$$

is equivalent to:

$$\underset{x,\,w}{\text{minimize}} \ \frac{1}{2}\|y - Ax\|_2^2 \quad \text{subject to} \quad x = g(z,w)$$

- network topology acts as a generic prior for “natural images”
- does not bias solution (untrained network adds little info)
- Non-convex and no guarantee that is (physically) feasible

Constrained weak deep prior

Replace w/ *weak* constraint:

$$\underset{x,\,w}{\text{minimize}}\ \frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda^2}{2}\|x - g(z,w)\|_2^2$$

and add *strong* handcrafted constraints (TV-norm):

$$\underset{x\in\mathcal{C},\,w}{\text{minimize}}\ \frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda^2}{2}\|x - g(z,w)\|_2^2$$

- imposes solution to remain in range generator w/ error proportional to λ
- strong constraints balanced w/ deep prior
- **strong handcrafted constraints guarantee (physically) feasibility**
- **does not sample posterior when training w/ *single* z**

Han, T., Lu, Y., Zhu, S.C. and Wu, Y.N., 2017, February. Alternating back-propagation for generator network. In Thirty-First AAAI Conference on Artificial Intelligence.

SLIM

Alternating back-propagation

A method for training generative models

- based on EM algorithm
- Does not need an extra network, e.g., recognition network in VAE, discriminator in GAN
- able to learn from incomplete or indirect data
  - This may prove difficult or less convenient for VAE and GAN
- **no need for x_i's**

SLIM

Setup

- training set of data vectors $\{Y_i, i = 1, ..., n\}$
- each Y_i has a corresponding Z_i (not observed)
- Probabilistic model: $[Y|Z, W] \sim p(Y|Z, W)$

$$Y = f(Z; W) + \epsilon,$$
$$Z \sim \mathrm{N}(0, I_d),\ \epsilon \sim \mathrm{N}(0, \sigma^2 I_D),\ d < D$$

Goal: Find parameters W that maximizes

$$p(Y; W) = \int p(Z)p(Y|Z, W)dZ \approx \sum_{i=1}^{n} \log p(Y_i; W) =$$

$$\sum_{i=1}^{n} \log \int p(Y_i, Z_i; W)dZ_i$$

Rubin, D.B. and Thayer, D.T., 1982. EM algorithms for ML factor analysis. Psychometrika, 47(1), pp.69-76.
Han, T., Lu, Y., Zhu, S.C. and Wu, Y.N., 2017, February. Alternating back-propagation for generator network. In Thirty-First AAAI Conference on Artificial Intelligence.

SLIM

EM algorithm to train generative models

$$
\begin{aligned}
&\frac{\partial}{\partial W} \log p(Y;W) = \frac{1}{p(Y;W)} \frac{\partial}{\partial W} p(Y;W) \\
&= \frac{1}{p(Y;W)} \int \frac{\partial}{\partial W} p(Z,Y;W) dZ = \frac{p(Z|Y;W)}{p(Z,Y;W)} \int \frac{\partial}{\partial W} p(Z,Y;W) dZ \\
&= \int p(Z|Y;W) \frac{1}{p(Z,Y;W)} \frac{\partial}{\partial W} p(Z,Y;W) dZ = \int p(Z|Y;W) \frac{\partial}{\partial W} \log p(Z,Y;W) dZ \\
&= \mathbb{E}_{p(Z|Y;W)} \left[\frac{\partial}{\partial W} \log p(Z,Y;W) \right]
\end{aligned}
$$

E-step: Approximate the expectation by drawing samples from $p(Z|Y,W)$ with

Stochastic Gradient Langevin Dynamics

Welling, M. and Teh, Y.W., 2011. Bayesian learning via stochastic gradient Langevin dynamics. In Proceedings of the 28th international conference on machine learning (ICML-11) (pp. 681-688).
https://www.ds3-datascience-polytechnique.fr/wp-content/uploads/2019/06/volkan-sampling-ds3-2019.pdf

Stochastic Gradient Langevin Dynamics

- Task: Given a target distribution $\mathrm{d}\mu = e^{-V(\mathbf{x})}\mathrm{d}\mathbf{x}$, generate samples from μ.
  - ▷ Most samples should be gathered around the minimum of V
  - ▷ We do not want convergence to the minimum
- Idea: Use Gradient Descent + noise

$$\mathbf{X}^{k+1} = \mathbf{X}^k - \beta_k \nabla V(\mathbf{X}^k) + \text{noise}$$

- Question: What amount of noise should we add so that $\mathbf{X}^\infty \sim \pi$?

Welling, M. and Teh, Y.W., 2011. Bayesian learning via stochastic gradient Langevin dynamics. In Proceedings of the 28th international conference on machine learning (ICML-11) (pp. 681-688).
https://www.ds3-datascience-polytechnique.fr/wp-content/uploads/2019/06/volkan-sampling-ds3-2019.pdf

SLIM

Large scale sampling

- A scalable framework: First-order sampling (assuming access to ∇V).

  Step 1. **Langevin Dynamics**

$$d\mathbf{X}_t = -\nabla V(\mathbf{X}_t)dt + \sqrt{2}d\mathbf{B}_t \quad \Rightarrow \quad \mathbf{X}_\infty \sim e^{-V}.$$

  Step 2. **Discretize**

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \beta_k \nabla V(\mathbf{x}^k) + \sqrt{2\beta_k}\boldsymbol{\xi}^k$$

  ▷ β_k step-size, $\boldsymbol{\xi}^k$ standard normal

  ▷ strong analogy to gradient descent method

Han, T., Lu, Y., Zhu, S.C. and Wu, Y.N., 2017, February. Alternating back-propagation for generator network. In Thirty-First AAAI Conference on Artificial Intelligence.

SLIM

Recall

Complete data log-likelihood

$$Y = f(Z; W) + \epsilon,$$
$$Z \sim \mathrm{N}(0, I_d),\ \epsilon \sim \mathrm{N}(0, \sigma^2 I_D),\ d < D$$

$$\Rightarrow$$

$$\log p(Y, Z; W) = \log\left[p(Z)p(Y|Z, W)\right]$$
$$= -\frac{1}{2\sigma^2}\|Y - f(Z; W)\|^2 - \frac{1}{2}\|Z\|^2 + \text{const.}$$

The posterior of Z can be written as (Bayes' rule)

$$p(Z|Y, W) = p(Y, Z; W)/p(Y; W) \propto \ p(Z)p(Y|Z, W)$$

Rubin, D.B. and Thayer, D.T., 1982. EM algorithms for ML factor analysis. Psychometrika, 47(1), pp.69-76.
Han, T., Lu, Y., Zhu, S.C. and Wu, Y.N., 2017, February. Alternating back-propagation for generator network. In Thirty-First AAAI Conference on Artificial Intelligence.

E-step: Langevin dynamics

$$p(Z|Y,W) = p(Y,Z;W)/p(Y;W) \propto \quad -\frac{1}{2\sigma^2}\|Y - f(Z;W)\|^2 - \frac{1}{2}\|Z\|^2$$

Samples from $p(Z|Y,W)$ are obtained from iterates:

$$Z_{\tau+1} = Z_\tau + sU_\tau + \frac{s^2}{2}\left[\frac{1}{\sigma^2}(Y - f(Z_\tau;W))\frac{\partial}{\partial Z}f(Z_\tau;W) - Z_\tau\right],$$

τ time step for the Langevin sampling
s step size
$U_\tau \sim \mathrm{N}(0, I_d)$

Rubin, D.B. and Thayer, D.T., 1982. EM algorithms for ML factor analysis. Psychometrika, 47(1), pp.69-76.
Han, T., Lu, Y., Zhu, S.C. and Wu, Y.N., 2017, February. Alternating back-propagation for generator network. In Thirty-First AAAI Conference on Artificial Intelligence.

SLIM

M-step: backpropagation

Now that we have samples from $p(Z|Y, W)$ we perform optimization of W

$$
\begin{aligned}
\mathbb{E}_{p(Z|Y;W)}\left[\frac{\partial}{\partial W}\log p(Z,Y;W)\right] &\approx \sum_{i=1}^{n}\frac{\partial}{\partial W}\log p(Y_i, Z_i; W) \\
&= -\sum_{i=1}^{n}\frac{\partial}{\partial W}\frac{1}{2\sigma^2}\|Y_i - f(Z_i;W)\|^2 \\
&= \sum_{i=1}^{n}\frac{1}{\sigma^2}(Y_i - f(Z_i;W))\frac{\partial}{\partial W}f(Z_i;W)
\end{aligned}
$$

After training/inversion: inference

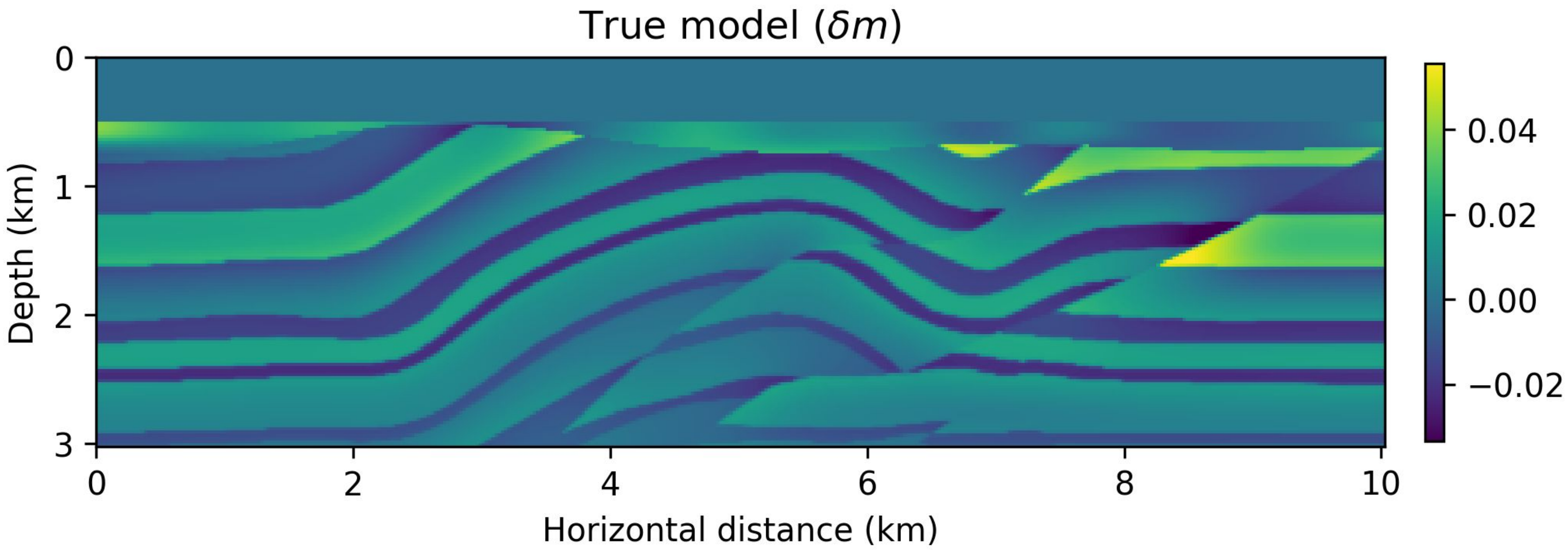
After training keep $w^* = w$ fixed

- draw samples from posterior via $x \sim g(z, w^*)$ with $z \sim N(0,1)$

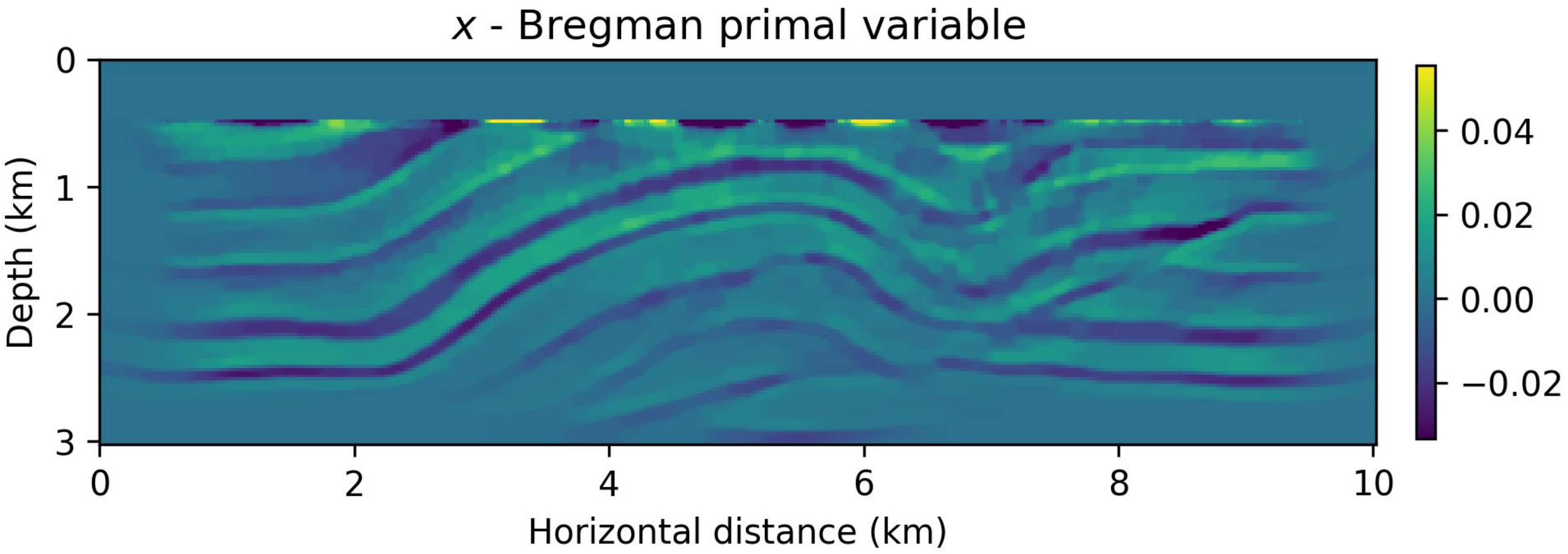
- derive statistical properties posterior simply via

$$\mathbb{E}_z \left\{ f\left(g(z, w^*)\right) \right\} \approx \frac{1}{K} \sum_{k=1}^{K} f\left(g(z_k, w^*)\right), \quad z_k \sim N(0,1)$$

True perturbation

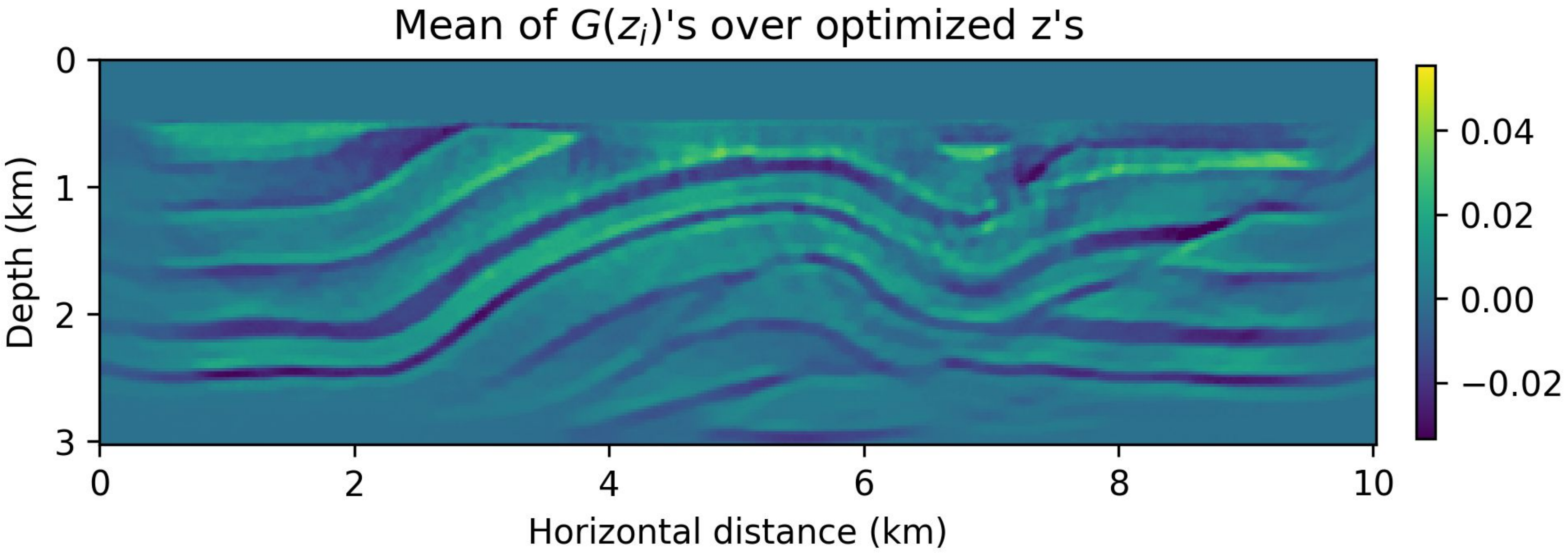


Slack variable in weak formulation



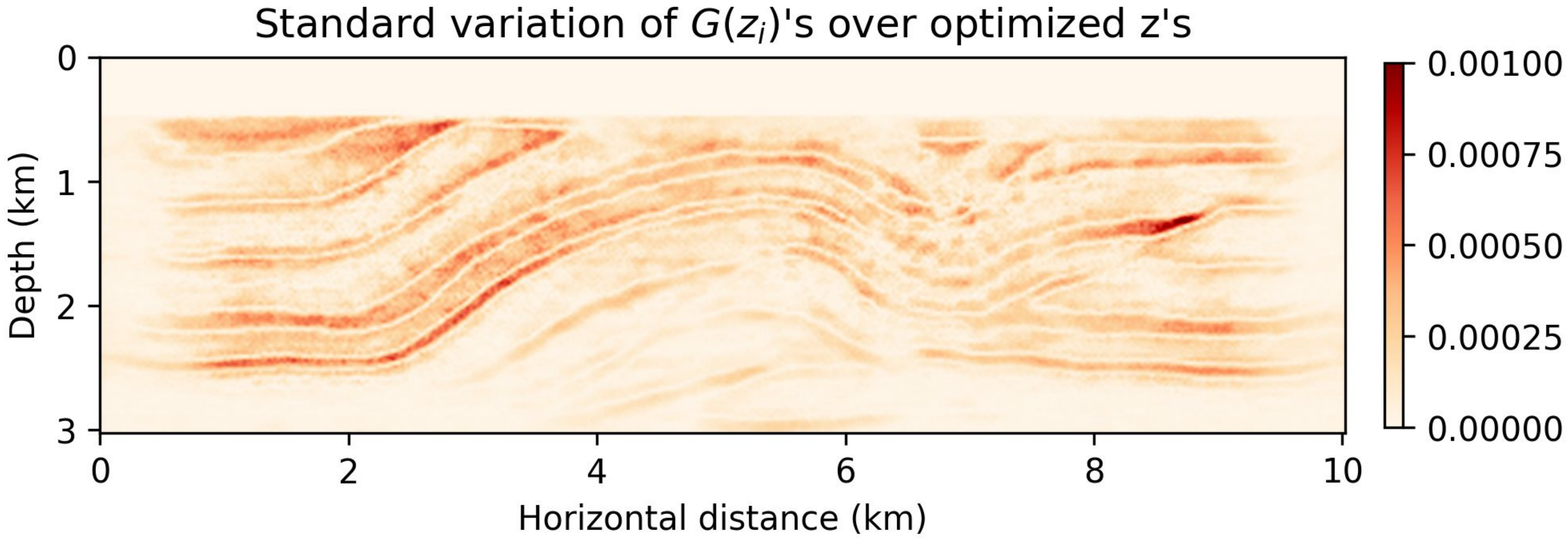
$$\underset{x \in \mathcal{C},\, w}{\text{minimize}} \ \frac{1}{2}\|y - Ax\|_2^2 + \frac{\lambda^2}{2}\|x - g(z, w)\|_2^2$$

Mean

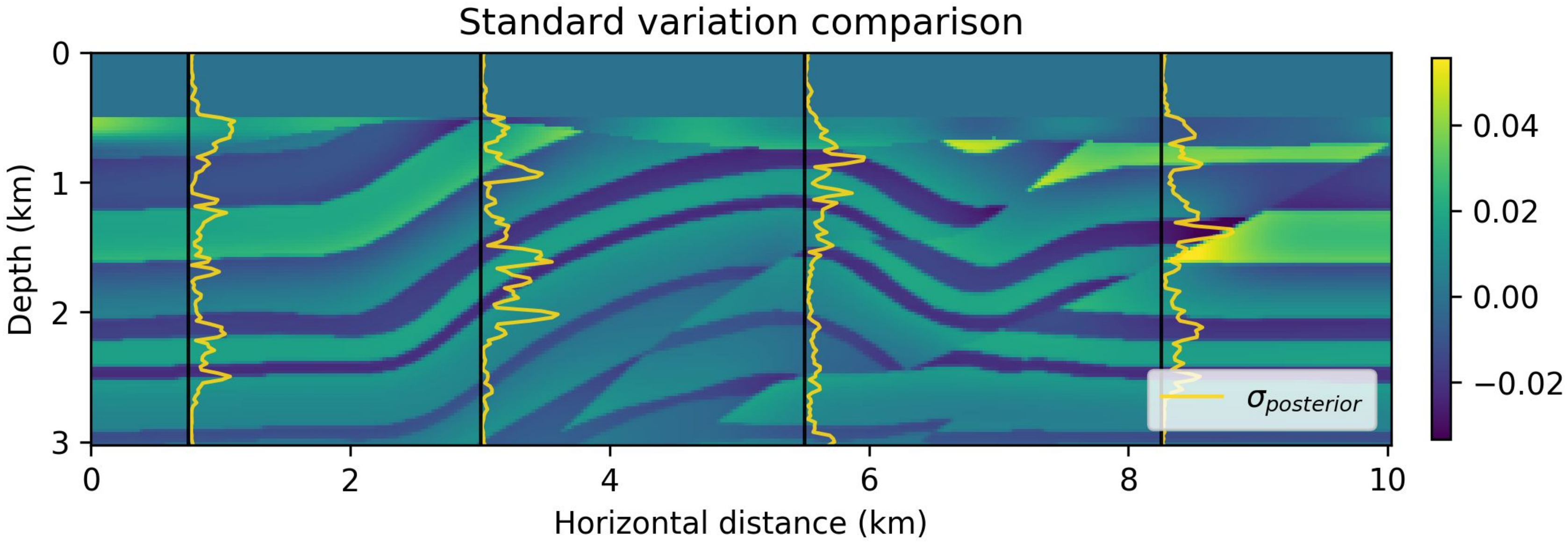


Mean of $G(z_i)$'s over optimized z's
Depth (km)
0
1
2
3
0
2
4
6
8
10
Horizontal distance (km)
0.04
0.02
0.00
−0.02

Pointwise variance



Pointwise variance: traces



Observations

Deep priors:

- regularize via their network topology
- made feasible in combination w/ strong handcrafted constraints
- reap information on the posterior during inversion

Computationally feasible fully-data driven framework that provides UQ

More theory & (nonlinear) examples needed…