SEG

WORKSHOP

FWI Beyond Structure Interpretation — Benefits and Challenges

10–12 February 2026 • Houston, TX, USA

image credit: doi.org/10.1190/tle38030179.1

When full waveform inversion meets gradient flows: an outlook on uncertainty quantification

FWI Beyond Structure Interpretation - Benefits and Challenges

Session: FWI Uncertainty

P. Marchner[1], R. Brossier[1], L. Métivier[1,2]

Thursday 12th February, 2026

[1]ISTerre, Univ. Grenoble Alpes, France

[1,2]LJK, CNRS, Univ. Grenoble Alpes, France

Outline

The uncertainty quantification problem

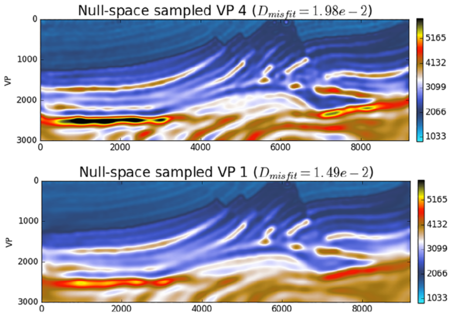
Gradient flows in probability space - a methodological framework for sampling

Application to FWI uncertainty quantification

Full Waveform Inversion

Given seismic observations d_{obs}, recover geological model $m(x)$ by minimizing a data misfit, e.g. $1/2\|d_{\text{cal}}[m] - d_{\text{obs}}\|_{L^2}^2$

Elastic Marmousi example (Liu and Peter (2020))



The inversion is ill-posed

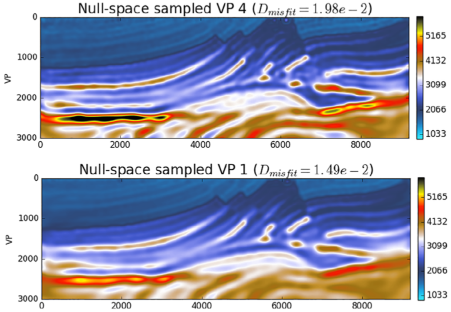
$\rightarrow$ non-unique reconstruction

The uncertainty quantification problem in FWI

Full Waveform Inversion

Given seismic observations d_{obs}, recover geological model $m(x)$ by minimizing a data misfit, e.g. $1/2\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{L^2}$

Elastic Marmousi example (Liu and Peter (2020))



The inversion is ill-posed

$\rightarrow$ non-unique reconstruction

Why ?

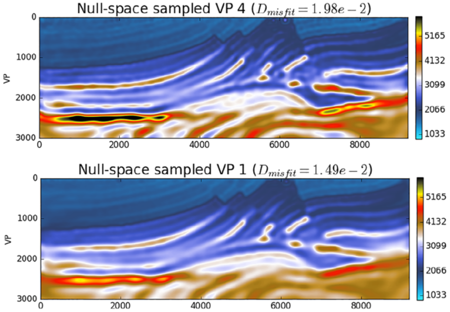
- noise in the observations d_{obs}
- physical modeling error (acoustic, elastic)
- cycle skipping
- lack of illumination in deep regions
- insufficient frequency content
- cross-talks between parameters
- ...

One inversion is not enough for decision-making

Full Waveform Inversion

Given seismic observations d_{obs}, recover geological model $m(\boldsymbol{x})$ by minimizing a data misfit, e.g. $1/2\|d_{\text{cal}}[m] - d_{\text{obs}}\|_{L^2}^2$

Elastic Marmousi example (Liu and Peter (2020))



The inversion is ill-posed

$\rightarrow$ non-unique reconstruction

Why ?

- noise in the observations d_{obs}
- physical modeling error (acoustic, elastic)
- cycle skipping
- lack of illumination in deep regions
- insufficient frequency content
- cross-talks between parameters
- ...

One inversion is not enough for decision-making

Uncertainty quantification : quantify the distribution of models consistent with the observations

The solution to the inverse problem is a probability distribution $\pi(m|d_{\text{obs}})$ (Tarantola and Valette (1982))

Bayesian formulation

Inverse problem : given noisy observations $d_{\text{obs}} \in \mathbb{R}^{n_{\text{obs}}}$, recover model $m(\boldsymbol{x}) \in \mathbb{R}^{d}$.

A common assumption : additive Gaussian noise

$$\boxed{d_{\text{obs}} = \mathcal{F}(m) + \eta, \quad \eta \sim \mathcal{N}(0, \Sigma)}$$

- $\Sigma \in \mathbb{R}^{n_{\text{obs}} \times n_{\text{obs}}}$: data covariance matrix $\rightarrow$ observation noise, model error, discretization, ...
- $\mathcal{F} : \mathbb{R}^{d} \rightarrow \mathbb{R}^{n_{\text{obs}}}$: forward map, $\mathcal{F}(m) := d_{\text{cal}}[m]$

Data likelihood distribution : $\pi(d_{\text{obs}}|m) \propto \exp\left(-\frac{1}{2}\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{\Sigma^{-1}}\right)$

Inverse problem : given noisy observations $d_{\text{obs}} \in \mathbb{R}^{n_{\text{obs}}}$, recover model $m(\boldsymbol{x}) \in \mathbb{R}^d$.
A common assumption : additive Gaussian noise

$$\boxed{d_{\text{obs}} = \mathcal{F}(m) + \eta, \quad \eta \sim \mathcal{N}(0, \Sigma)}$$

- $\Sigma \in \mathbb{R}^{n_{\text{obs}} \times n_{\text{obs}}}$: data covariance matrix $\rightarrow$ observation noise, model error, discretization, ...
- $\mathcal{F} : \mathbb{R}^d \rightarrow \mathbb{R}^{n_{\text{obs}}}$: forward map, $\mathcal{F}(m) := d_{\text{cal}}[m]$

Data likelihood distribution : $\pi(d_{\text{obs}}|m) \propto \exp\left(-\frac{1}{2}\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{\Sigma^{-1}}\right)$

Bayesian statistics

Given some prior knowledge $\pi_{\text{prior}}(m)$, find the posterior distribution (Kaipio and Somersalo (2006))

Bayes theorem for inverse problems

$$\pi_{\text{post}}(m) := \pi(m|d_{\text{obs}}) = \frac{\pi(d_{\text{obs}}|m)\pi_{\text{prior}}(m)}{\pi(d_{\text{obs}})}, \quad \pi(d_{\text{obs}}) \neq 0$$

Gaussian prior $\pi_{\text{prior}}(m) \sim \mathcal{N}(m_0, \Sigma_0) \Rightarrow \pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2}\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{\Sigma^{-1}} - \frac{1}{2}\|m - m_0\|^2_{\Sigma_0^{-1}}\right)$
If $\mathcal{F}$ is linear, the posterior is also Gaussian with explicitly known mean and covariance

Sampling the posterior in FWI

The challenges in FWI :

- Nonlinear forward map $\mathcal{F} \Rightarrow$ non-Gaussian posterior
- High-dimensional model space ($d \gg 1$)
- Expensive evaluation of $\mathcal{F}$: one sample $\approx$ one wave propagation PDE solve

The sampling task : generate samples $\{m^{(i)}\}_{i=1}^{N_s}$ of models from $\pi_{\text{post}}(m)$

Direct sampling of the posterior is computationally intractable !

Sampling the posterior in FWI

The challenges in FWI :

- Nonlinear forward map $\mathcal{F} \Rightarrow$ non-Gaussian posterior
- High-dimensional model space ($d \gg 1$)
- Expensive evaluation of $\mathcal{F}$: one sample $\approx$ one wave propagation PDE solve

The sampling task : generate samples $\{m^{(i)}\}_{i=1}^{N_s}$ of models from $\pi_{\text{post}}(m)$

Direct sampling of the posterior is computationally intractable !

Brief state of the art

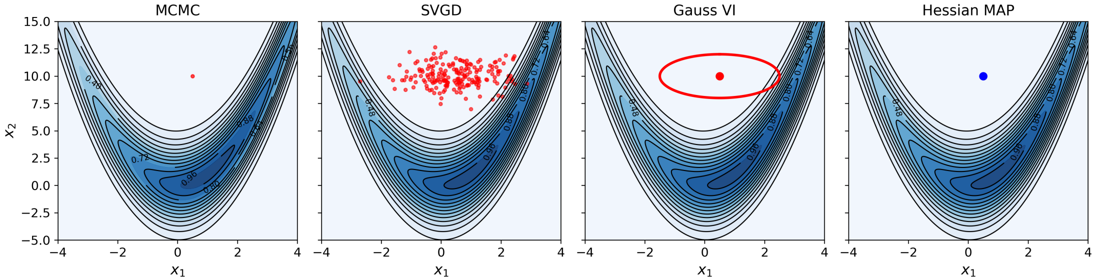
- Small to moderate size problems : MCMC, HMC (Gebraad et al. (2020)), RJMCMC (Sen and Biswas (2017))
- Large-scale : Hessian at MAP (Bui-Thanh et al. (2013)), GVI (Zhang et al. (2023)), Norm. flows (Orozco et al. (2025))

Markov Chain Monte Carlo	variational inference SVGD, normalizing flows	Gaussian propagation EnKF, Gauss VI	optimization, local analysis at maximum *a posteriori*

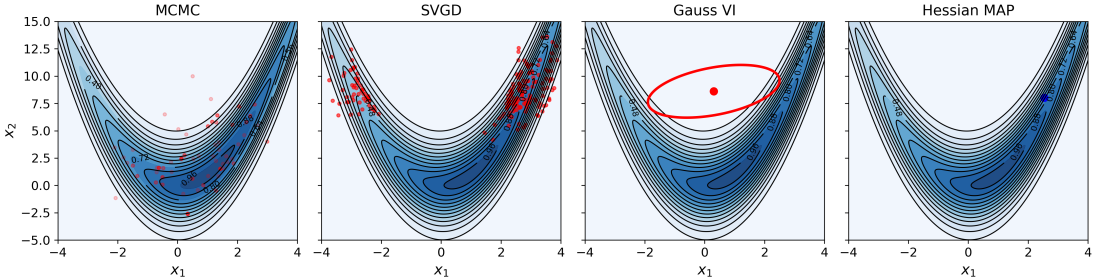
costly, global exploration ← ? → cheaper, local exploration

Rosenbrock density example

Rosenbrock function adapted to sampling : $(x_1^\star, x_2^\star) = (1, 1)$ is the maximum a posteriori (MAP)

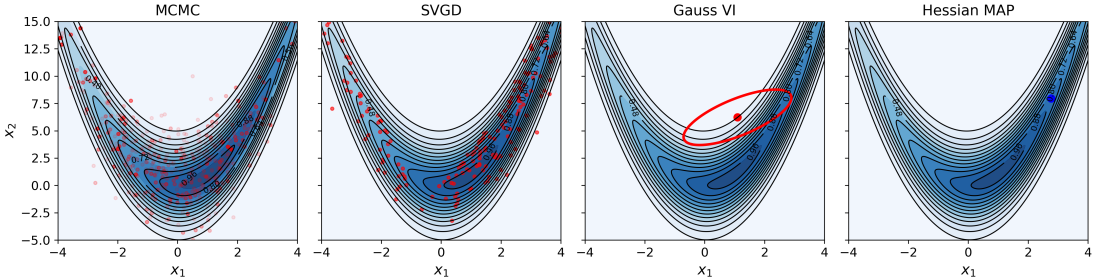


Rosenbrock function adapted to sampling : $(x_1^\star, x_2^\star) = (1, 1)$ is the maximum a posteriori (MAP)

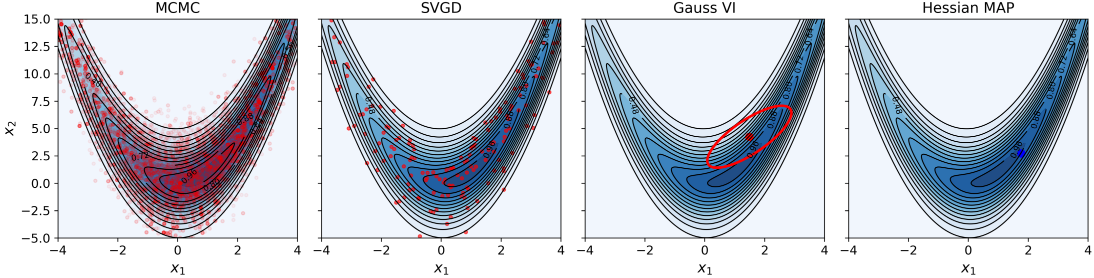


Rosenbrock density example

Rosenbrock function adapted to sampling : $(x_1^\star, x_2^\star) = (1, 1)$ is the maximum a posteriori (MAP)

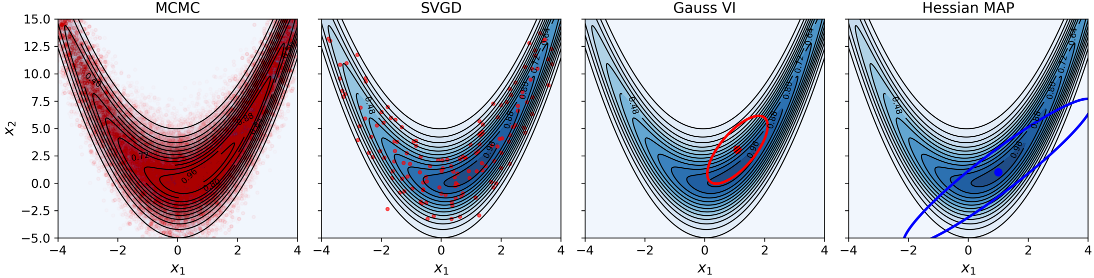


Rosenbrock function adapted to sampling : $(x_1^\star, x_2^\star) = (1, 1)$ is the maximum a posteriori (MAP)



Rosenbrock density example

Rosenbrock function adapted to sampling : $(x_1^\star, x_2^\star) = (1,1)$ is the maximum a posteriori (MAP)



$\#$ forward map calls : Random walk MCMC $>$ SVGD $>$ Gaussian VI $>$ Hessian at MAP

Questions

- How to design efficient methods : exploration vs accuracy w.r.t. π_{post} vs cost ?
- What statistical information do we want from the posterior ?

Outline

The uncertainty quantification problem

Gradient flows in probability space - a methodological framework for sampling

Application to FWI uncertainty quantification

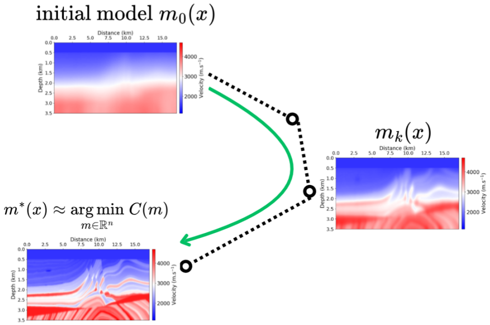
In FWI, we minimize $C(m) = \|d_{\text{cal}}[m] - d_{\text{obs}}\|_{L^2}^2$ iteratively

The usual gradient descent scheme

$$m_{k+1} = m_k - \alpha_k P(m_k)\nabla C(m_k), \quad \text{preconditioner } P,$$

is an explicit Euler discretization of a gradient flow in $\mathbb{R}^d$

$$\frac{dm}{d\tau} = -\nabla_P C(m) := -P(m)\nabla C(m), \quad m_0 = m_{\text{init}}$$

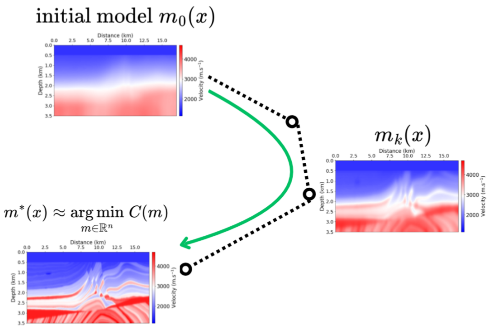


In FWI, we minimize $C(m) = \|d_{\mathrm{cal}}[m] - d_{\mathrm{obs}}\|^2_{L^2}$ iteratively
The usual gradient descent scheme

$$m_{k+1} = m_k - \alpha_k P(m_k) \nabla C(m_k), \quad \text{preconditioner } P,$$

is an explicit Euler discretization of a gradient flow in $\mathbb{R}^d$

$$\frac{dm}{d\tau} = -\nabla_P C(m) := -P(m) \nabla C(m), \quad m_0 = m_{\mathrm{init}}$$

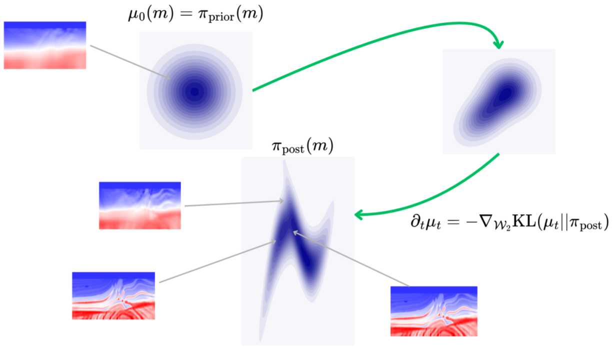


We need three ingredients

1. a misfit function $C(m)$
2. a space equipped with an inner product $(\mathbb{R}^d, \langle \cdot, \cdot \rangle_{P^{-1}})$ which defines the gradient ∇_P
3. a numerical scheme

Can we lift this picture on probability densities ?

Probabilistic FWI as gradient flow in probability space

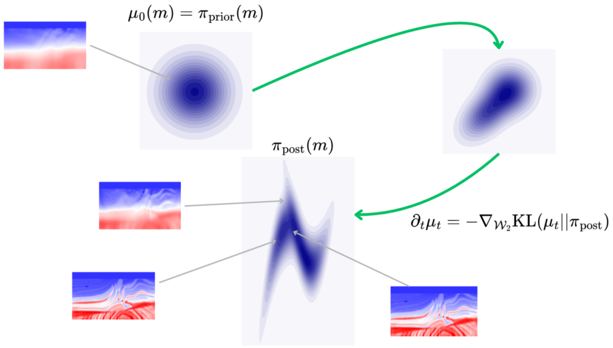


We need three ingredients

1. a misfit functional to minimize : relative information between μ and π_{post}

$$\boxed{\text{KL}(\mu || \pi_{\text{post}}) = \int \mu(m) \log \left(\frac{\mu(m)}{\pi_{\text{post}}(m)} \right) dm}$$

Probabilistic FWI as gradient flow in probability space

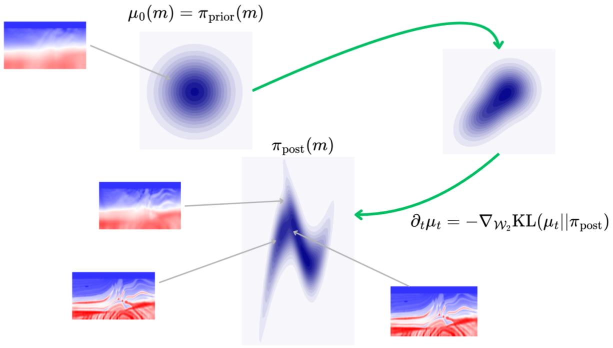


We need three ingredients

1. a misfit functional to minimize : relative information between μ and π_{post}

$$\boxed{\text{KL}(\mu||\pi_{\text{post}}) = \int \mu(m) \log \left(\frac{\mu(m)}{\pi_{\text{post}}(m)} \right) dm}$$

2. a geometry on densities, induced by the 2-Wasserstein metric, which defines the gradient $\nabla_{\mathcal{W}_2}$ (Santambrogio (2015); Engquist et al. (2016))



We need three ingredients

1. a misfit functional to minimize : relative information between μ and π_{post}

$$\boxed{\text{KL}(\mu||\pi_{\text{post}}) = \int \mu(m) \log \left(\frac{\mu(m)}{\pi_{\text{post}}(m)} \right) dm}$$

2. a geometry on densities, induced by the 2-Wasserstein metric, which defines the gradient $\nabla_{\mathcal{W}_2}$ (Santambrogio (2015); Engquist et al. (2016))
3. a numerical scheme to follow the flow
   - how to represent the density μ ?
   - how to discretize the flow in "time" ?

Idea. Gradient flows on probabilities describe how *mass is transported* to minimize an objective functional.

Gradient flow of the KL divergence with 2-Wasserstein metric

$$\partial_\tau \mu_\tau = -\nabla_{\mathcal{W}_2} \mathrm{KL}(\mu_\tau \| \pi_{\text{post}}) \iff \partial_\tau \mu_\tau = \nabla \cdot (\mu_\tau \nabla C) + \Delta \mu_\tau \quad \text{(Fokker–Planck equation)}$$

Convergence guarantee : For a convex misfit C (or log-concave π_{post}), $\mu_\tau \to \pi_{\text{post}}$ as $\tau \to \infty$.

SEISCOPE

Idea. Gradient flows on probabilities describe how *mass is transported* to minimize an objective functional.

Gradient flow of the KL divergence with 2-Wasserstein metric

$$\partial_\tau \mu_\tau = -\nabla_{W_2} \mathrm{KL}(\mu_\tau \| \pi_{\text{post}}) \iff \partial_\tau \mu_\tau = \nabla \cdot (\mu_\tau \nabla C) + \Delta \mu_\tau \quad \text{(Fokker–Planck equation)}$$

Convergence guarantee : For a convex misfit C (or log-concave π_{post}), $\mu_\tau \to \pi_{\text{post}}$ as $\tau \to \infty$.

Numerical schemes $\Rightarrow$ sampling algorithms

Particle representation

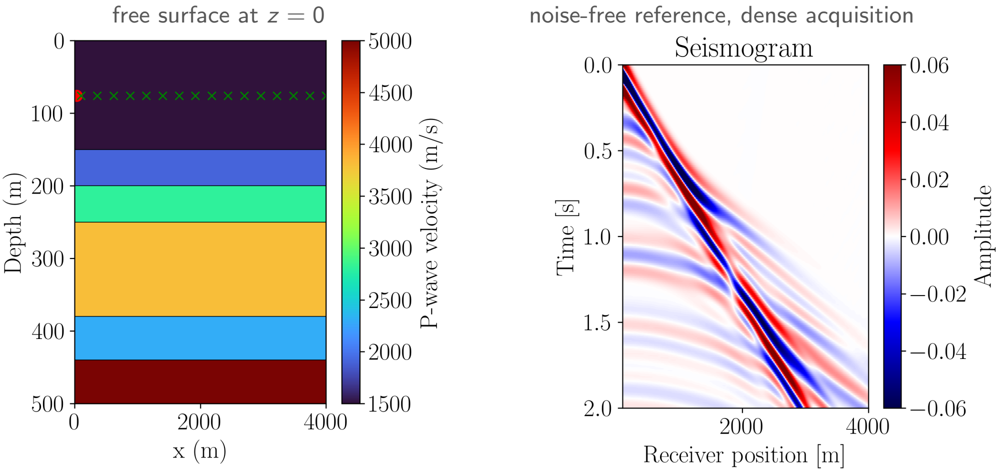
- Langevin dynamics
- SVGD

Parametric representation

- Gaussian variational inference
- Normalizing flows

In both standard and probabilistic FWI, the dynamics of the flow is determined by the geometry.

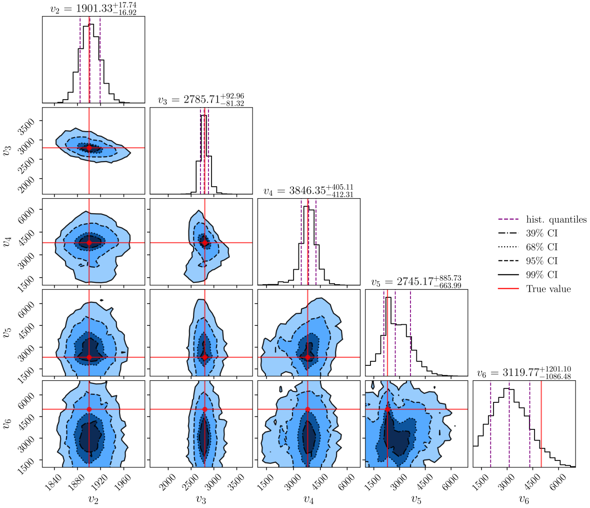
A synthetic reference benchmark



2D layer model: $m = (v_2, v_3, v_4, v_5, v_6)^T$. Forward map (acoustic) : reflectivity method (Mallick and Frazer (1987))

$v_1 = 1500$ m/s fixed, single Ricker source $f_0 = 5$ Hz, $T = 2.0$ sec, 16 receivers, 20 dB SNR

$$\pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2}\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{\Sigma^{-1}} - \frac{1}{2}\|m - m_0\|^2_{\Sigma_0^{-1}}\right)$$



Broad Gaussian prior $\pi_{\text{prior}}(m) \sim \mathcal{N}(m_0, \Sigma_0)$

$m_0 = (2, 2.5, 2.5, 3, 3)^T$ km/s

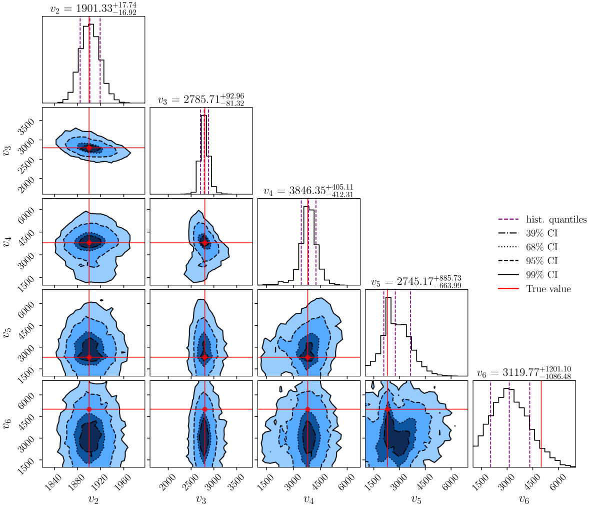
$\Sigma_0 = \text{diag}(1.^2, 1.5^2, 1.5^2, 1.5^2, 1.5^2)$ km/s

Σ : diag covariance, 8:1 energy balance with prior

Reference posterior $\rightarrow \mathcal{O}(10^5)$ forward calls

Reference posterior distribution

$$\pi_{\text{post}}(m) \propto \exp\left(-\frac{1}{2}\|d_{\text{cal}}[m] - d_{\text{obs}}\|^2_{\Sigma^{-1}} - \frac{1}{2}\|m - m_0\|^2_{\Sigma_0^{-1}}\right)$$



Broad Gaussian prior $\pi_{\text{prior}}(m) \sim \mathcal{N}(m_0, \Sigma_0)$

$m_0 = (2, 2.5, 2.5, 3, 3)^T$ km/s

$\Sigma_0 = \text{diag}(1.^2, 1.5^2, 1.5^2, 1.5^2, 1.5^2)$ km/s

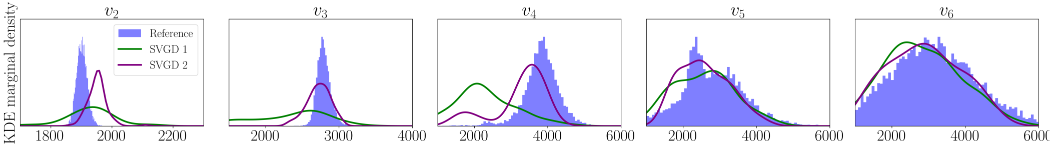
Σ : diag covariance, 8:1 energy balance with prior

Reference posterior $\rightarrow \mathcal{O}(10^5)$ forward calls

Posterior features

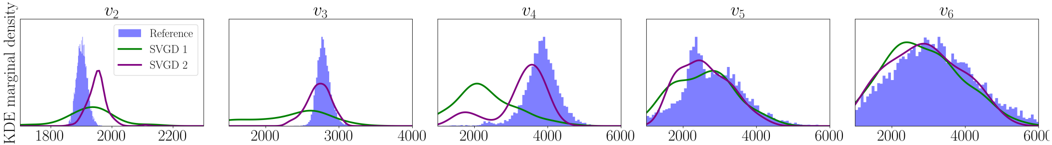
- non-Gaussian, anisotropic
- multi-modal (v_4, v_5 marginals)
- data information weakens with depth
  $\rightarrow$ higher uncertainty

SVGD - $\mu_{\text{init}} = \mathcal{N}(m_0, \Sigma_0/2)$, $\mu_{\text{init}} = \mathcal{N}(m_0, \Sigma_0/6)$ with 100 particles



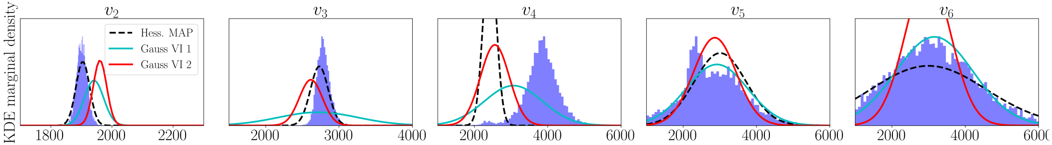
Number of forward calls : $2 \times n_{\text{it}} \times n_{\text{particles}} \rightarrow \mathcal{O}(10^4)$

SVGD - $\mu_{\text{init}} = \mathcal{N}(m_0, \Sigma_0/2)$, $\mu_{\text{init}} = \mathcal{N}(m_0, \Sigma_0/6)$ with 100 particles



Number of forward calls : $2 \times n_{\text{it}} \times n_{\text{particles}} \rightarrow \mathcal{O}(10^4)$

Gaussian VI and Hessian at MAP - $\mu_{\text{init}} = \mathcal{N}(m_0, \Sigma_0/2)$, $\mu_{\text{init}} = \mathcal{N}(m_0, \Sigma_0/6)$



Number of forward calls :

- Gauss VI : $2 \times n_{\text{it}} \times n_{\text{quad}} \rightarrow \mathcal{O}(10^3)$
- Hessian MAP $\rightarrow \mathcal{O}(10^2)$

Standard schemes struggle on complex posteriors,
$\rightarrow$ cycle-skipping

How to scale to larger problems ? EnKF as a Gaussian variational update with dimension reduction

Gaussian variational inference

$\min_{\overline{m}, C} \mathrm{KL}(\mu || \pi_{\text{post}})$, with $\mu(m) \sim \mathcal{N}(\overline{m}, C)$

+ mean update $\overline{m}$ only
- full covariance matrix
- 1st order scheme

Ensemble Kalman Filter (EnKF) - Evensen (2003)

$\mu(m) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(m - m^{(i)})$

- N particles updates $m^{(i)}$, $i = \{1, \ldots, N\}$
+ low-rank covariance from particles
+ 2nd order, Gauss-Newton scheme

How to scale to larger problems ? EnKF as a Gaussian variational update with dimension reduction

Gaussian variational inference

$\min_{\overline{m}, C} \mathrm{KL}(\mu || \pi_{\text{post}})$, with $\mu(m) \sim \mathcal{N}(\overline{m}, C)$

+ mean update $\overline{m}$ only
- full covariance matrix
- 1st order scheme

Ensemble Kalman Filter (EnKF) - Evensen (2003)

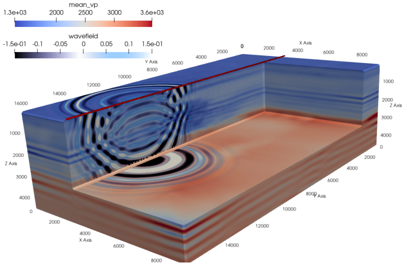
$\mu(m) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(m - m^{(i)})$

- N particles updates $m^{(i)}$, $i = \{1, \ldots, N\}$
+ low-rank covariance from particles
+ 2nd order, Gauss-Newton scheme

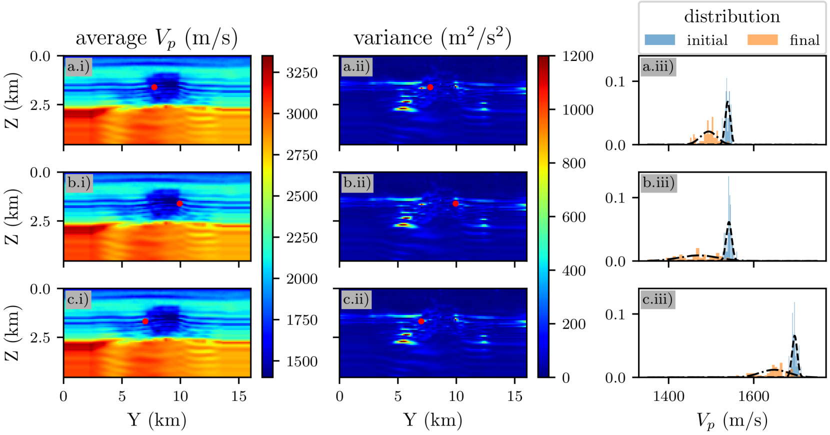
Application to field data - Hoffmann et al. (2024)

- 4-component Ocean Bottom Cable device
- VTI acoustic model, pressure component for the inversion : V_p inversion only
- 2048 reciprocal sources, 2.5-5Hz frequency band
- Sequential Bayes updates by batch of 16 shots $\rightarrow$ 3 ℓ-BFGS + 1 EnKF update for each batch

Wavefield at $t = 3.3$ s on mean FWI-EnKF model



Initial ensemble in a local valley to avoid cycle-skipping, $N = 50$ particles enough for Gaussian uncertainty



Final mean and variance at 3 points of interest, section at $x = 2.95$ km

EnKF-FWI : reasonably cheap UQ where richer inference methods are (yet) too expensive

Summary

- Uncertainty quantification (UQ) for FWI requires sampling a non-Gaussian posterior
- Gradient flow theory provides a rigorous framework to understand and design UQ algorithms
- The framework guides novel physics-aware algorithms with convergence guarantees
  $\rightarrow$ Gaussian mixture VI, EnKF without collapse, ℓ-BFGS like preconditioning, ...
- Scaling to large problems requires dimension reduction and careful preconditioning

Summary

- Uncertainty quantification (UQ) for FWI requires sampling a non-Gaussian posterior
- Gradient flow theory provides a rigorous framework to understand and design UQ algorithms
- The framework guides novel physics-aware algorithms with convergence guarantees
  $\rightarrow$ Gaussian mixture VI, EnKF without collapse, ℓ-BFGS like preconditioning, ...
- Scaling to large problems requires dimension reduction and careful preconditioning

Takeaways

- UQ is fundamentally about moving probability mass wisely
- Gradient flows connects sampling algorithms with the physics and difficulties of FWI

Thank you for your attention and thanks to

- All SEISCOPE sponsors : AKERBP, DUG, EXXONMOBIL, GEOLINKS, JGI, PETROBRAS, SHEARWATER, SHELL, TOTALENERGIES and VIRIDIEN
- IDRIS, TGCC and CINES, French national computing centers
- GRICAD, Grenoble computing center
- SWAN, Hewlett Packard Enterprise (HPE) Cray XC System
- All SEISCOPE project members

Questions ?

philippe.marchner@univ-grenoble-alpes.fr

Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G. (2013). A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523.

Engquist, B., Froese, B. D., and Yang., Y. (2016). Optimal transport for seismic full waveform inversion. *Communications in Mathematical Sciences*, 14(8):2309–2330.

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367.

Gebraad, L., Boehm, C., and Fichtner, A. (2020). Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo. *Journal of Geophysical Research: Solid Earth*, 125(3).

Hoffmann, A., Brossier, R., Métivier, L., and Tarayoun, A. (2024). Local uncertainty quantification for 3-D time-domain full-waveform inversion with ensemble Kalman filters: application to a North Sea OBC data set. *Geophysical Journal International*, 237(3):1353–1383.

Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media.

Liu, Q. and Peter, D. (2020). Square-root variable metric-based nullspace shuttle: A characterization of the nonuniqueness in elastic full-waveform inversion. *Journal of Geophysical Research: Solid Earth*, 125(2).

Mallick, S. and Frazer, L. N. (1987). Practical aspects of reflectivity modeling. *Geophysics*, 52(10):1355–1364.

Orozco, R., Siahkoohi, A., Louboutin, M., and Herrmann, F. J. (2025). ASPIRE: iterative amortized posterior inference for Bayesian inverse problems. *Inverse Problems*, 41(4):045001.

Santambrogio, F. (2015). *Optimal transport for applied mathematicians*. Birkäuser, Basel.

Sen, M. K. and Biswas, R. (2017). Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm. *Geophysics*, 82(3):R119–R134.

Tarantola, A. and Valette, B. (1982). Inverse problems= quest for information. *Journal of geophysics*, 50(1):159–170.

Zhang, X., Lomas, A., Zhou, M., Zheng, Y., and Curtis, A. (2023). 3-D Bayesian variational full waveform inversion. *Geophysical Journal International*, 234(1):546–561.