Multiscale uncertainty quantification for post-stack seismic inversion with wavelet flows

Gabrio Rizzuti, Ivan Vasconcelos

Abstract

Point-estimate statistics for inverse problems, such as the maximum a-posteriori estimator, require the solution of a problem that is typically ill-conditioned for seismic imaging applications, with important implications in terms of computational complexity. Ill-conditioning is arguably an even bigger challenge for uncertainty quantification, which aims at a more comprehensive characterisation of the posterior distribution. One classical strategy to assuage these issues is the multiscale approach, where the original problem is broken down into a hierarchical sequence of sub-problems with increasing computational complexity. Each sub-problem describes scale-dependent features of the overall solution for any chosen scale-dependent decomposition. This gives rise to an efficient iterative method that progressively builds a solution from "coarse" to "fine" scales. We propose to leverage recent developments in machine-learning-based variational inference for uncertainty quantification that uses a wavelet-based generative model of the posterior distribution. The architectural design is based on a normalising flow that generates the scales of a sample sequentially and conditionally based on the coarser scales. As a first application of this framework, we study post-stack seismic inversion here.

Novelty: We apply a multiscale neural network based on the combination of normalising flows and wavelet decomposition for estimating uncertainty in inverse problems

Key aspects: The multiscale architecture of the network allows a scale-by-scale sequential estimation of the uncertainty, thus reducing ill-conditioning and further increasing computation efficiency for both training and inference at progressively coarser scales

Multiscale uncertainty quantification for post-stack seismic inversion with wavelet flows

Introduction

Acoustic post-stack seismic inversion aims at characterising the *acoustic impedance* of the medium from stacked images or data. In 2D, the acoustic impedance unknowns are generally discretised on a time/common-midpoint coordinate grid (t, x) and denoted by $\boldsymbol{x} \in R^{n_t \times n_x}$. Data $\boldsymbol{y}$ indicate stacked seismic traces, conventionally obtained from migration techniques, and are denoted by $\boldsymbol{y} \in R^{n_t \times n_x}$.

According to the Bayes' rule, the *posterior* distribution of $\boldsymbol{x}$ knowing $\boldsymbol{y}$ is given by $p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{y})$, where $p(\boldsymbol{y}|\boldsymbol{x})$ and $p(\boldsymbol{x})$ are, respectively, the *likelihood* and the *prior* distributions. We assume a (known) linear data-generation model:

$$-\log p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{2\sigma^2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \cdots, \qquad \boldsymbol{A} = \boldsymbol{W}\boldsymbol{D}_t \tag{1}$$

The operator $\boldsymbol{W}$ denotes a convolution in time with an estimated wavelet, while $\boldsymbol{D}_t$ is the time derivative. The symbol $\|\cdot\|$ indicates the least-squares norm. Finally, the data noise standard deviation, tied to $p(\boldsymbol{y}|\boldsymbol{x})$, is σ. Additional terms not relevant to this paper are indicated by "⋯".

We assume that the prior distribution is differentiable and available in analytical form. For simplicity, however, we focus on Gaussian priors of the form:

$$-\log p(\boldsymbol{x}) = \frac{1}{2\tau_1^2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 + \frac{1}{2\tau_2^2}\|\boldsymbol{B}\boldsymbol{x}\|^2 + \cdots. \tag{2}$$

Here, $\boldsymbol{x}_0$ is a background model. The weighting matrix $\boldsymbol{B} = \left(\boldsymbol{I} - \boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}\right)\boldsymbol{\nabla}_{tx}$ promotes models whose gradient field $\boldsymbol{\nabla}_{tx}\boldsymbol{x}$ aligns with some (given) prior structural information. The scalar quantities τ_1 and τ_2 govern the relative weighting of the regularisation terms in Equation 2.

The problem object of this paper is the characterisation of the posterior distribution for some given data $\boldsymbol{y}$ (a single sample). The desired solution consists of generating random samples $\boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{y})$, from which any statistical moment can then be approximated. While the problem in Equations 1 and 2 admits a straightforward analytical solution in 2D, it is useful as a benchmark for comparing various uncertainty quantification (UQ) techniques.

Prior art and contributions

The most classical methods for UQ are based on Markov-chain Monte Carlo sampling (MCMC). While the history of MCMC in geophysics is deep, its application remains computationally challenging, especially for 3D non-linear imaging problems.

Variational inference is an alternative approach where the target posterior distribution is approximated by a member of a parameterised family of distributions, which is optimal in some sense. For example, particle-based variational inference parametrises the candidate distribution directly in terms of model samples: this is the case for the Stein Variational Gradient Descent method (SVGD). Notable applications can be found in Zhang et al. (2023) and Izzatullah et al. (2023). Alternatively, the distribution can be implicitly represented by a generative model based on neural networks. The class of normalising flows (NF, Kobyzev et al., 2021, Orozco et al., 2023b) is of particular interest here, since it allows the analytical computation of its own output density function. Seismic applications include those by Rizzuti et al. (2020), Zhang and Curtis (2021), Siahkoohi et al. (2023), Orozco et al. (2023a).

In this work, we focus on multiscale UQ techniques. The primary motivation consists in the need to effectively extend the scope of UQ to large-sized ill-conditioned inverse problems, both for the sake of computationally efficiency in UQ and to accommodate scale-dependent UQ for multiple seismic applications, ranging from post-stack inversion to full-waveform inversion (FWI). We discuss the novel application of a particular normalising flow architecture, originally presented in Yu et al. (2020) for generative modelling, based on a wavelet-based scale decomposition of the posterior distribution. The

network allows scale-wise sequential training with evident computational benefits. Contrary to the work of Yu et al. (2020), we are not tackling a classical generative modelling application in the sense that training samples $\boldsymbol{x}\sim p(\boldsymbol{x}|\boldsymbol{y})$ are not available, but the density function analytical form is. Furthermore, the motivations for the wavelet decomposition are different. In Yu et al. (2020) the NFs associated to different scales can be trained independently and in parallel to scale up NF-based generative modelling. In our case, the scale-dependent distributions are rather approximated sequentially from coarse to fine levels for the purpose of reducing the ill-conditioning of the inverse problem under examination.

Multiscale normalising flow framework

We consider the family of candidate distributions $\boldsymbol{x}\sim p_{\boldsymbol{\theta}}(\boldsymbol{x}) \approx p(\boldsymbol{x}|\boldsymbol{y})$ represented by the transport maps $T_{\boldsymbol{\theta}}: Z \to X$, e.g.

$$\boldsymbol{x}\sim p_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \Leftrightarrow \quad \boldsymbol{x} = T_{\boldsymbol{\theta}}(\boldsymbol{z}), \boldsymbol{z}\sim p(\boldsymbol{z}). \tag{3}$$

The latent-space variables $\boldsymbol{z}$ are drawn from a simple distribution (e.g. a standard Gaussian). The transport maps are parametrised by the unknowns $\boldsymbol{\theta}$, which correspond to the parameters of a neural network. We then seek to minimise the Kullback-Leibler divergence:

$$\min_{\boldsymbol{\theta}} \mathrm{KL}(p_{\boldsymbol{\theta}}(\boldsymbol{x})\| p(\boldsymbol{x}|\boldsymbol{y})) = E_{\boldsymbol{z}\sim p(\boldsymbol{z})} - \log p(T_{\boldsymbol{\theta}}(\boldsymbol{z})|\boldsymbol{y}) + \log p_{\boldsymbol{\theta}}\big(T_{\boldsymbol{\theta}}(\boldsymbol{z})\big) + \cdots. \tag{4}$$

The first term in Equation 4 is the average reconstruction error, while the second term penalises mode-collapsing to the maximum a-posteriori estimate (MAP). For general transport maps, calculating and differentiating this term is intractable. However, special neural architectures such as normalising flows are designed to provide this quantity exactly (Kobyzev et al., 2021).

In this paper, we focus on normalising flows. Several competing architectural choices are available in the literature, but we adopt a multiscale structure based on the Haar wavelet decomposition (Yu et al., 2020), dubbed Wavelet flow (see Figure 1). The low- and high-resolution components of the (2D) wavelet transform $\boldsymbol{W}^k$ applied to an image $\boldsymbol{x}^k \in R^{n_t/2^k \times n_x/2^k}$ are denoted as $\boldsymbol{x}^{k+1} \in R^{n_t/2^{k+1} \times n_x/2^{k+1}}$ and $\tilde{\boldsymbol{x}}^{k+1} \in R^{n_t/2^{k+1} \times n_x/2^{k+1} \times 3}$, respectively. The superscript k indicates a certain resolution level, with $\boldsymbol{x}^0 = \boldsymbol{x}$. Due to orthogonality, this implies:

$$\boldsymbol{W}^k\boldsymbol{x}^k = \big(\boldsymbol{x}^{k+1}, \tilde{\boldsymbol{x}}^{k+1}\big) \to p\big(\boldsymbol{x}^k\big) = p\big(\boldsymbol{x}^{k+1}, \tilde{\boldsymbol{x}}^{k+1}\big) = p\big(\tilde{\boldsymbol{x}}^{k+1}|\boldsymbol{x}^{k+1}\big)p\big(\boldsymbol{x}^{k+1}\big). \tag{5}$$

By applying recursively the wavelet transform up to a certain scale K, we obtain the following probabilistic chain:

$$p(\boldsymbol{x}^0) = p(\tilde{\boldsymbol{x}}^1|\boldsymbol{x}^1)p(\tilde{\boldsymbol{x}}^2|\boldsymbol{x}^2)\cdots p(\tilde{\boldsymbol{x}}^K|\boldsymbol{x}^K)p(\boldsymbol{x}^K), \tag{6}$$

with, again, $\boldsymbol{x}^0 = \boldsymbol{x}$. Yu et al. (2020) leverage the scale-wise decomposition in Equation 6 by training individual conditional NFs to approximate the conditional densities $p_{\boldsymbol{\theta}^k}\big(\tilde{\boldsymbol{x}}^k|\boldsymbol{x}^k\big) \approx p\big(\tilde{\boldsymbol{x}}^k|\boldsymbol{x}^k\big)$ and a final NF for the coarsest scale distribution $p_{\boldsymbol{\theta}^K}(\boldsymbol{x}^K) \approx p(\boldsymbol{x}^K)$.

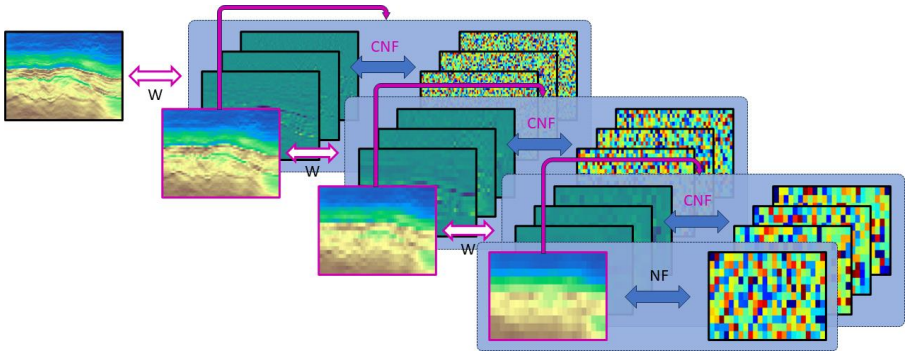

Figure 1 Wavelet flow architecture. The wavelet transform **W** decomposes an image on several scale levels, the image distribution at each scale is learned by a dedicated NF.

Benchmark results: Volve field data

We consider 2D post-stack seismic inversion for a selected section of the Volve field data set (equinor.com/energy/volve-data-sharing). Pre-processing (data selection, velocity, and wavelet estimation) is carried out after Izzatullah et al. (2023). In Figure 2, we show the post-stack data and a smoothed background acoustic impedance.

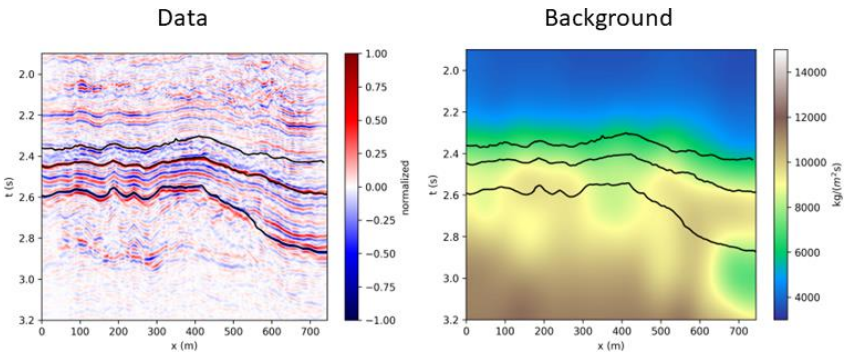

Figure 2 A 2D section of the Volve data set together with background acoustic impedance

We perform UQ according to several methods. The analytic method provides ground-truth results, which can be computed due to the simplifying assumptions in Equation 1 and Equation 2 (e.g., Gaussian likelihood and prior). The NF-based method was described in the previous section. We also compare these results with the particle-based SVGD method, which shares many similarities with our proposed scheme but as an optimisation approach presents challenges with initialisation and numerical cost scaling. The first-order statistics, e.g. the conditional mean and (point-wise) standard deviation, for these three methods are displayed in Figure 3.

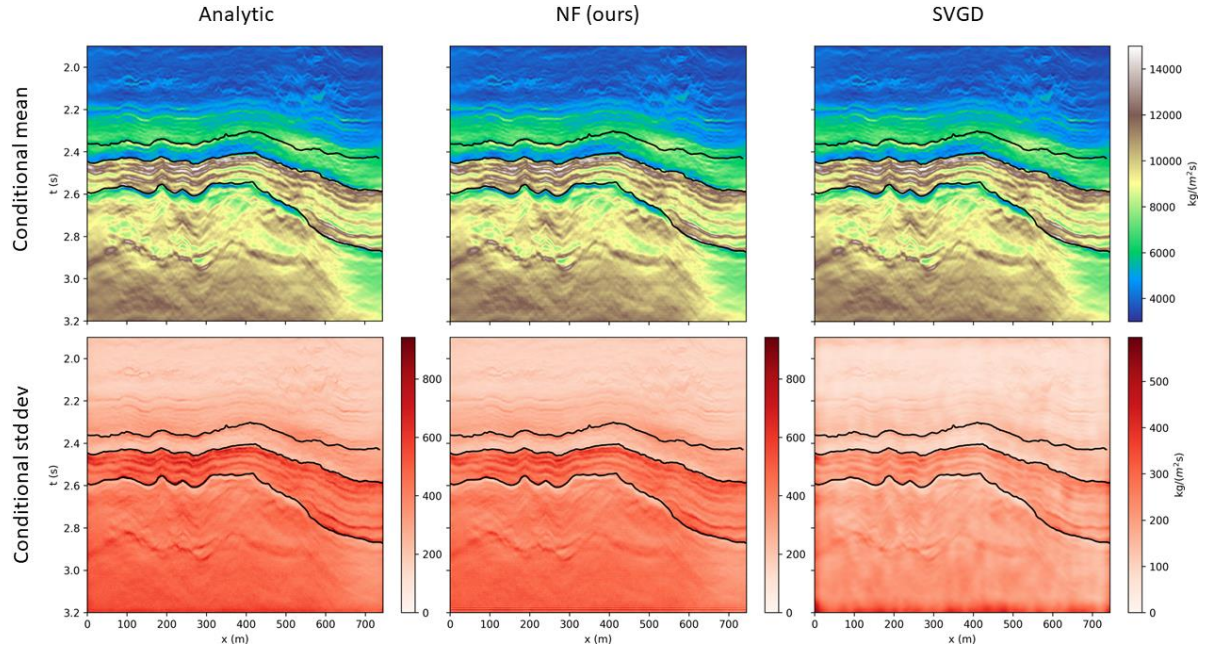

Figure 3 Comparison of first-order statistics for different UQ methods

Discussion and conclusions

All three methods compared in Figure 3 produce near-identical conditional means. This is due to how both the NF and SVGD methods are initialised (that is, as perturbations around the pre-computed MAP). Regarding the conditional standard deviation, the NF-based method is quantitatively close to the ground truth, with only some light underestimation. The SVGD-based UQ, on the other hand, is heavily biased

due to the limited number of particles (equal to 128 in this case; see the standard deviation scaling in Figure 3). As noted in Zhang et al. (2023), stochastic variants of SVGD might reduce the bias. We remark that NF- and SVGD-based UQ both seek to minimise the KL divergence with transport maps. Furthermore, the output of the NF can be considered a "particle" and the batch size used in the training phase can be equated to the particle number in SVGD, but we argue that NF-based UQ is intrinsically more efficient than SVGD due to the underlying functional representation of the particle distribution (as opposed to the explicit particle representation in SVGD).

The multiscale nature of the proposed method is highlighted in Figure 4, where the first-order statistics are computed for random samples produced at selected scales. We note that a multiscale approach can be included in SVGD, in principle, but this does not fundamentally resolve its memory-complexity limitations due to the particle number, while it is a by-product of our NF framework.

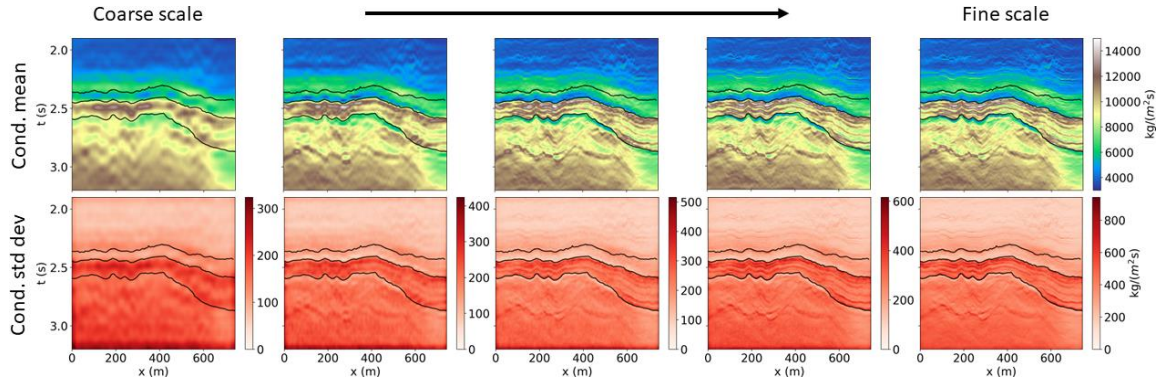

Figure 4 UQ for several resolution scales as a by-product of the multiscale NF-based method

The framework described in the previous sections can be applied to more general settings. For example it can deal with non-linear forward operators and/or non-Gaussian priors. Aside from looking at more complex imaging applications (such as tomography or FWI), future work might also explore the role of learned wavelet decomposition and loop-unrolling techniques to include problem-specific inductive bias directly in the network architecture.

References

Izzatullah, M., Alkhalifah, T., Romero, J., Corrales, M., Luiken, N. and Ravasi, M. [2023] Posterior sampling with CNN-based, Plug-and-Play regularisation with applications to Post-Stack Seismic Inversion. *Geophysics*, 1–73.

Kobyzev, I., Prince, S.J.D. and Brubaker, M.A. [2021] Normalising Flows: An Introduction and Review of Current Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**(11), 3964–3979.

Orozco, R., Louboutin, M. and Herrmann, F.J. [2023a] Fast neural FWI with amortised uncertainty quantification. In: *International Meeting for Applied Geoscience and Energy*.

Orozco, R., Witte, P., Louboutin, M., Siahkoohi, A., Rizzuti, G., Peters, B. and Herrmann, F.J. [2023b] InvertibleNetworks.jl: A Julia package for scalable normalizing flows. ArXiv:2312.13480 [cs].

Rizzuti, G., Siahkoohi, A., Witte, P.A. and Herrmann, F.J. [2020] Parameterising uncertainty by deep invertible networks: An application to reservoir characterisation. In: *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 1541–1545.

Siahkoohi, A., Rizzuti, G., Orozco, R. and Herrmann, F.J. [2023] Reliable amortised variational inference with physics-based latent distribution correction. *Geophysics*, **88**(3).

Yu, J.J., Derpanis, K.G. and Brubaker, M.A. [2020] Wavelet Flow: Fast Training of High Resolution Normalising Flows. In: *Advances in Neural Information Processing Systems*, 33. 6184–6196.

Zhang, X. and Curtis, A. [2021] Bayesian Geophysical Inversion Using Invertible Neural Networks. *Journal of Geophysical Research: Solid Earth*, **126**(7).

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.