Efficient Seismic Inversion via Diffusion Posterior Sampling with Probability Flow ODE

Chaoshun Hu, Nizar Chemingui and Raafat Abdul Alim*
TGS

Summary

We present a seismic inversion framework that combines poststack or post-fwi inversion with pre-trained diffusion generative models. Although diffusion models provide a powerful prior for noisy inverse problems, diffusion posterior sampling is often constrained by the high cost of repeated likelihood-score evaluations through Tweedie-based gradients. To address this limitation, we replace the reverse-time stochastic differential equation with the deterministic probability flow ODE (PFODE). This formulation enables larger time steps, faster inference, and better scalability while maintaining high inversion quality. Results on the Marmousi model and the Volve field dataset show that the PFODE approach preserves structural continuity, improves subsurface recovery, and lowers computational cost. Overall, the proposed method provides an efficient, robust, and high-fidelity framework for seismic reconstruction.

Introduction

Full waveform inversion (FWI) has become a standard tool for high-resolution seismic imaging by minimizing the misfit between observed and modeled wavefields (Tarantola,1984). Through iterative updates of subsurface parameters, FWI can in principle recover fine-scale velocity structures beyond the limits of conventional migration methods. However, despite decades of development, FWI remains highly nonlinear and ill-posed. The inversion process is sensitive to acquisition geometry, limited offset coverage, restricted bandwidth. These limitations frequently lead to cycle-skipping, slow convergence, or entrapment in local minima.

Various strategies have been proposed to mitigate these issues, including adaptive waveform inversion (Warner and Guasch, 2016), time shift and dynamic matching algorithm (Luo and Schuster, 1991; Zhang et al., 2018; Mao et al, 2020), optimal transport formulations (Enquist and Yang, 2022), and gradient decomposition (Tang et al, 2013) and preconditioning techniques (Yong et al,, 2010; Yao et al. 2019). While effective in many scenarios, these approaches remain fundamentally deterministic. They typically provide a single best-fit model and do not explicitly characterize model uncertainty. Furthermore, they do not incorporate learned geological prior information beyond what is implicitly encoded in the starting model or regularization term.

In recent years, generative models from the machine learning community have opened new directions for solving inverse problems in a Bayesian framework (Dimakis et al, 2022; Sahlstrom and Tarvainen, 2023). Rather than directly minimizing a deterministic objective, Bayesian approaches aim to characterize the posterior distribution of subsurface models conditioned on the observed data. Generative models can serve as powerful implicit priors that encode realistic geological structures learned from training datasets. Among these models, diffusion models have demonstrated remarkable success in high-dimensional generation tasks and have recently attracted attention in seismic imaging and inverse problems (Song et al., 2021a; Chung et al., 2022; Shi et al., 2024).

Diffusion models (Ho et al., 2020; Song et al., 2021b) begin by defining a forward noising process that gradually transforms clean data into Gaussian noise and a learned reverse process that reconstructs data from noise. Once trained, the reverse process can be interpreted as a learned score function of the data distribution. Diffusion posterior sampling (DPS) leverages this score in combination with a physics-based likelihood gradient to produce samples consistent with measurements (Chung et al. 2022). This framework is particularly appealing for seismic imaging because it decouples prior learning from forward modeling and allows training-free guidance using the wave equation (Wang et al., 2023; Shen et al., 2025).

However, most DPS implementations (Chung et al., 2022; Song et al., 2023) formulate the reverse process as a stochastic differential equation (SDE). Although mathematically well founded, SDE-based sampling poses several practical challenges for seismic applications. First, stable integration typically requires small time steps especially when subsurface structure is complex, increasing the number of function evaluations. Second, stochastic sampling introduces variability across runs, which can be undesirable in production environments requiring reproducibility when variances have large range. Third, each likelihood evaluation involves forward and adjoint wave equation computations, making repeated backpropagation computationally costly. The recent InverseBench study (Zheng et al., 2025) also notes that all plug-and-play diffusion-prior (PnPDP) methods are computationally less efficient than conventional baselines. These factors collectively limit the scalability of SDE-based DPS for large-scale seismic imaging.

An alternative, but mathematically equivalent, formulation is given by the probability flow ordinary differential equation (Song et al., 2023). The PFODE shares the same marginal distributions as the reverse SDE but removes the stochastic noise term, yielding a deterministic trajectory in model space. This deterministic nature enables larger integration step sizes, improved numerical stability, and reduced sampling variance. From a computational perspective, the PFODE can significantly reduce the number of required likelihood evaluations, which is particularly important for wave-equation-based inverse problems.

In this work, we develop a PFODE-based diffusion posterior sampling framework tailored to seismic imaging. We derive a deterministic formulation for variance-preserving diffusion models and extend it to inverse problems through training-free guidance driven by physics-based gradients. We further employ a dedicated high-order DPM solver for diffusion ODEs with guaranteed convergence order (Lu et al., 2022). Benchmark tests on synthetic and field data show that the PFODE approach delivers high-quality reconstructions while substantially improving computational efficiency.

Theory - From Diffusion to Probability Flow ODE

Diffusion models are characterized by forward and reverse processes. In the forward process, an image is progressively transformed into Gaussian noise over the time interval from 0 to T. Conversely, the reverse process evolves from T back to 0, reconstructing the image from the noise. Here we use x_t to denote the state of the system at time t; The forward process can be described by the following SDE (Song et al, 2021):

$$dx_t = f(t)x_t dt + g(t)d\omega \quad (1)$$

where ω is standard Brownian motion. If we consider a variance-preserving (VP) SDE, the drift term $f(t)$ is defined as $f(t) = -\frac{1}{2}\beta(t)$ and the diffusion coefficients is $g(t) = \sqrt{\beta(t)}$, where $\beta(t)$ is the noise schedule over time t.

The reverse process can be formulated similarly to the forward process and is described by the following SDE:

$$dx_t = \left[f(t)x_t - g_t^2 \nabla_{x_t} \log p_t(x_t)\right]dt + g(t)d\bar{\omega}_t \quad (2)$$

where $\bar{\omega}_t$ satisfies a standard Wiener process when time t reverses from T to 0, $p_t(x_t)$ is the time-dependent data distribution, and $\nabla_{x_t} \log p_t(x_t)$ represents the score function at each time step.

Equations (1) and (2) are inherently stochastic, leading to variable intermediate trajectories and non-reproducible results. Upon discretization, the SDE step size is constrained by the randomness of the Wiener process. As a result, larger step sizes can cause divergence, particularly in high-dimensional spaces. However, for all diffusion processes, there exists a corresponding deterministic process whose trajectories share the same marginal probability $p_t(x_t)$ as the SDE. Song et al. proved that the probability flow ODE for Eq (1) can be expressed as:

$$dx_t = \left[f(t)x_t - \frac{1}{2}g_t^2 \nabla_{x_t} \log p_t(x_t)\right]dt \quad (3)$$

Theory – Training-free Guidance

In order to use probability flow ODE model for solving inverse problems using posterior sampling, one can solve Eq. (3) with replacing $\nabla_{x_t} \log p_t(x_t)$ with its problem specific counterpart, $\nabla_{x_t} \log p_t(x_t|y)$, where $p_t(x_t|y)$ represents the PDF of the noise variable at time step t given the measurement $y = Ax + n$ regarded as the result of applying a forward operator A to the target x, corrupted by additive Gaussian noise n, as follows:

$$dx_t = \left[f(t)x_t - \frac{1}{2}g_t^2 \nabla_{x_t} \log p_t(x_t|\, y)\right]dt \quad (4)$$

where the score function is replaced by conditional score function.

To calculate the problem-specific score function, Bayes' rule can be used,

$$\nabla_{x_t} \log p_t(x_t|y) = \nabla_{x_t} \log p_t(x_t) + \nabla_{x_t} \log p_t(y|x_t) \quad (5)$$

where $\nabla_{x_t} \log p_t(x_t)$ can be approximated using score estimation networks trained on the original data, as a part of the diffusion model training process. The second term (guidance term) in RHS Eq. (5) adapts the sampling procedure to produce samples that are consistent with the given measurement. Considering this term is intractable at state $\boldsymbol{x_t}$, we leverage Tweedie formula to approximate the expectation as follows,

$$\nabla_{x_t} \log p_t(y|x_t) \approx \nabla_{x_t} \log p_t\big(y|\hat{x}_0(x_t)\big) \quad (6)$$

where $\hat{\boldsymbol{x}}_{\boldsymbol{0}}$ can be approximated with $\alpha(t) = 1 - \beta(t)$ and $\bar{\alpha}(t) = \prod_{i=1}^{t} \alpha(t)$ as follows,

$$\hat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}(t)}}\Big(x_t + \big(1 - \bar{\alpha}(t)\big)s_\theta(x_t, t)\Big) \quad (7).$$

In recent diffusion models for FWI implementations, $\nabla_{x_t} \log p_t(y|x_t)$ is computed with FWI gradients with training-free guidance,

$$\nabla_{x_t} \log p_t(y|x_t) \approx -\nabla_{x_t} \|y - A\hat{x}_0(x_t)\|_2^2 \quad (8)$$

However, with such training-free guidance, the direct relationship in Eq. (8) does not always hold with $\boldsymbol{\iota_2}$ loss function. At initial phase, the loss remains relatively high but it oscillates for larger t. This is partially because training-free guidance is more sensitive to the adversarial gradient which is a significant challenge for neural networks, which refers to minimal perturbations deliberately applied to inputs that can induce disproportionate alterations in the model's output (Shen et al., 2024).

Probability Flow ODE for Seismic Imaging

Theory – PF-ODE based diffusion posterior sampling (PFODE-DPS)

A key challenge is the evaluation of the noise-perturbed likelihood score $\nabla_{x_t} \log p_t(d|x_t)$. Existing approximations, including DPS, PGDM, and OT-ODE (Chung et al., 2022; Song et al., 2023; Pokle et al., 2023), improve inversion quality but remain computationally expensive because each sampling step requires gradients through the pre-trained generative model. Their stochastic formulation also limits reproducibility. To overcome these limitations, we adopt the probability flow ODE (PFODE) and develop a deterministic diffusion posterior sampling framework that combines physics-based likelihood guidance with a diffusion prior. Under the Bayesian formulation, the dynamics are decomposed into likelihood and prior terms as follows,

$$dx_t = \underbrace{\left[f(t)x_t - \frac{1}{2}g(t)^2\nabla_{x_t}\log p_x(x_t)\right]dt}_{prior\ term} \underbrace{-\frac{1}{2}g(t)^2\nabla_{x_t}\log p(y|x_t)dt}_{likelihood\ term} \quad (9)$$

The prior term is advanced with a dedicated first-order diffusion solver, reducing the number of sampling steps to about 20-100, compared with roughly 1000 in the forward diffusion process. The corresponding prior update is given by the equation

$$x_{t-1}^{prior} = \frac{\tilde{\alpha}_{t-1}}{\tilde{\alpha}_t}x_t + \tilde{\alpha}_{t-1}\tilde{\beta}_t\left(\frac{\tilde{\beta}_t}{\tilde{\alpha}_t} - \frac{\tilde{\beta}_{t-1}}{\tilde{\alpha}_{t-1}}\right)s_\theta(x_t) \quad (10)$$

where $\tilde{\alpha}_t = \sqrt{\bar{\alpha}_t}$ and $\tilde{\beta}_t = \sqrt{1-\bar{\alpha}_t}$.

Figure 1 shows the trajectory comparison between the 1D conventional Denoising Diffusion Probabilistic Model (DDPM) and the proposed PFODE trajectories. Compared to DDPM, PFODE gives deterministic results at the data side.

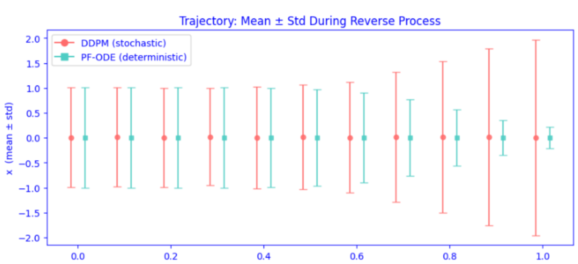

Figure 1, trajectory comparison of mean ± std between DDPM and PFODE during reverse process.

Examples

To evaluate the proposed PFODE method for seismic imaging, we test it on the Marmousi model and the Volve field dataset. Figure 2 shows unconditional samples generated by the pre-trained model from clean velocity models. In training, we use 64×64 patches extracted from public benchmark models, including Marmousi, Hess VTI, SEG Overthrust, Amoco Topography, and SEAM. These results illustrate the generative capability of the pre-trained model prior to conditioning on seismic data.

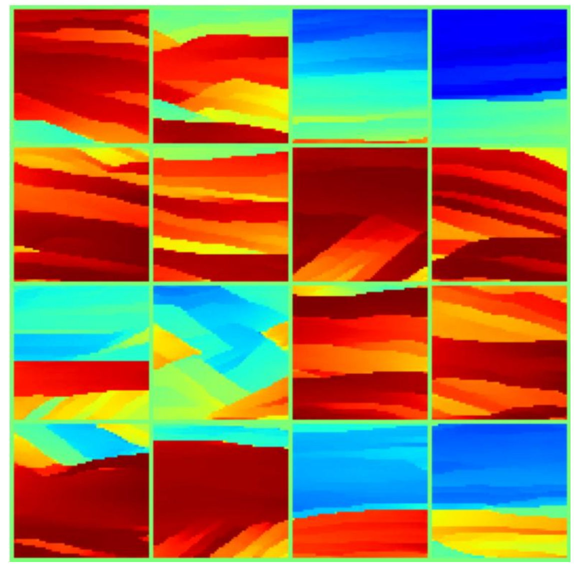

Figure 2, unconditional sampling from pretrained model

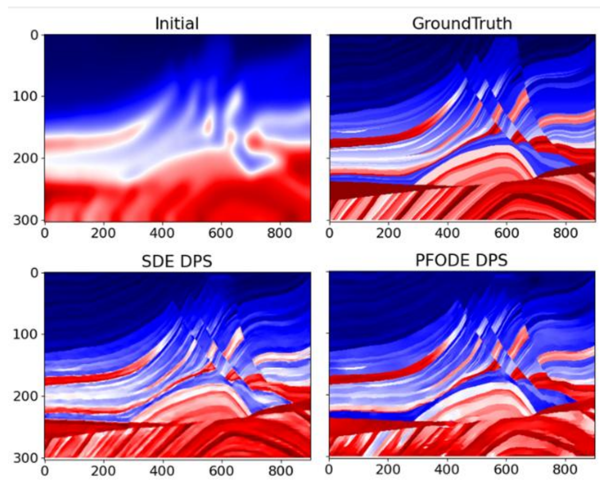

Figure 3. Seismic inversion result comparison using diffusion posterior sampling as a prior for Marmousi data

In our first benchmark test, we generate zero-offset data using a born modeling code from the velocity perturbation and then compute the corresponding postmigration image with the adjoint operator in time domain. This stack image-velocity pair enables the application of both SDE DPS and PFODE-DPS posterior sampling. In the proposed PFODE-

DPS implementation, we also incorporate second-order gradient information through Hessian. Figure 3 compares the inversion results for the 2D Marmousi model. Relative to SDE-based DPS, PFODE sampling achieves comparable reconstruction quality with significantly fewer function evaluations. Using 25 steps, the proposed approach provides about a 7× speedup over the two-stage SDE-based DPS process. The PFODE DPS method achieves a SSIM value at 0.8950 while the SDE-based method has a SSIM value at 0.8612.

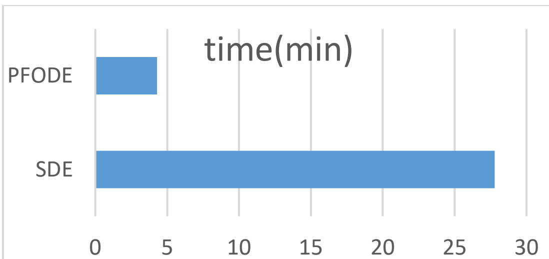

Figure 4. Runtime comparison between the SDE-based method and the proposed two-stage PFODE-based method

We also test the method on a 2d seismic section data from the public Volve field dataset. The original seismic data is preprocessed with a series of steps developed by Kirch in 2011 from the ST10010ZC11 survey in Volve field. We extract the seismic data and migration velocity section at inline# 9999. We simply scale the velocity model to the acoustic impedance, convert to reflectivity and then born modeling is used to generate stack data for inversion. The field-data velocity inversion results in Figure 5 demonstrate stable imaging performance, reduced oscillatory artifacts, and improved continuity of subsurface features. The ability to use larger integration steps enhances computational efficiency, making the approach more practical for production-scale workflows.

Conclusion

We propose a probability flow ODE formulation for diffusion posterior sampling in seismic imaging. By replacing stochastic reverse diffusion with a deterministic trajectory, the method improves reproducibility, stability, and computational efficiency relative to SDE-based approaches. From a geophysical perspective, it provides a practical way to combine learned geological priors with wave-equation-based inversion, yielding reconstructions that are both data-consistent and geologically plausible. Results on Marmousi and the Volve field dataset show that the PFODE approach preserves structural continuity and improves subsurface recovery at lower computational cost. These features make it a promising scalable Bayesian tool for seismic inversion and subsurface characterization.

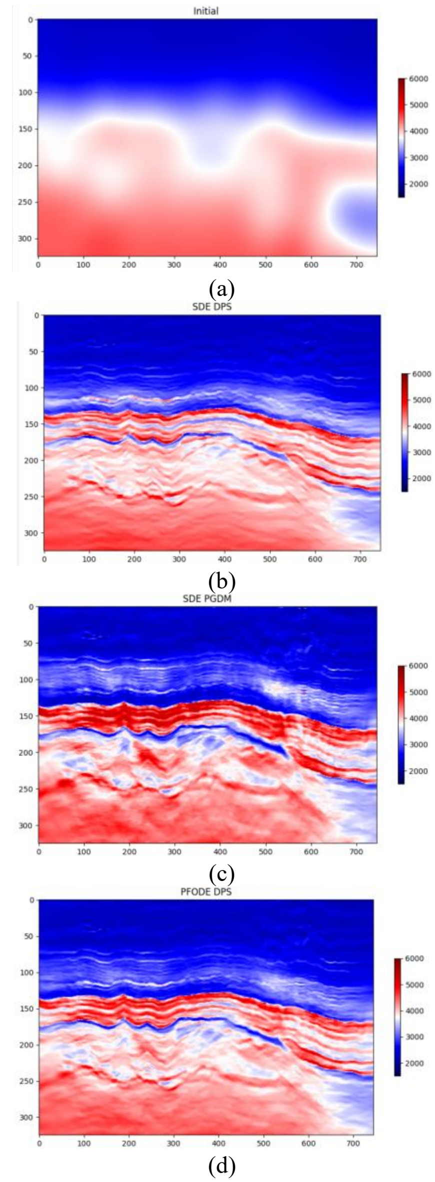

Figure 5, seismic inversion result comparison using diffusion posterior sampling as a prior for Volve dataset.

Acknowledgement

We thank TGS management for their permission to publish this result. We also thank Equinor for releasing the public dataset from the Volve field. The first author also thanks Yutong He for her teaching CMU 10-799 Diffusion & Flow matching course and Matteo Ravasi's diffusion work at KAUST.