Solving Inverse Problems with Stochastic Interpolants: Self-Consistent Generative Modeling from Corrupted Data

Anonymous Author(s)
Affiliation
Address
email

Abstract

Transport-based methods have emerged as a leading paradigm for building generative models from large datasets. However, in many scientific and engineering domains, clean data samples are often unavailable: instead, we only observe corrupted measurements obtained through a noisy, ill-conditioned forward map. We introduce a novel framework for *inverse generative modeling* that learns to generate clean data using only these corrupted observations. Our approach leverages stochastic interpolants to construct a self-consistent training procedure: we iteratively transport corrupted observations to clean data samples, then verify consistency by passing the generated samples back through the forward map to match the original observation distribution. This bypasses the need for clean training data while maintaining theoretical guarantees. The resulting method is (i) computationally efficient compared to variational alternatives, and (ii) highly flexible, handling arbitrary nonlinear forward models with only black-box access. We demonstrate superior performance on a variety of inverse problems arising in imaging applications.

1 Introduction

Transport-based methods have become the primary approach for training high-quality generative models with large amounts of data. In many scientific and engineering applications, however, direct access to the data $x \sim \pi$ of interest is unavailable. Instead, we only observe corrupted measurements y through a *forward* map $y = \mathcal{F}(x)$ that is typically noisy and ill-conditioned. Examples include medical imaging, where we observe tomographic projections of internal structures, astronomical observations affected by atmospheric distortion, and other measurement processes that introduce noise and information loss [Tarantola, 2005].

In this work, we introduce a framework for inverse generative modeling that learns to generate clean data x using only these corrupted observations y and simulation of the forward process $\mathcal{F}$. Our approach leverages stochastic interpolants (SI) [Albergo and Vanden-Eijnden, 2023, Albergo et al., 2023], building on the broader family of transport-based methods including flow matching [Lipman et al., 2022] and rectified flows [Liu et al., 2023]. We construct a self-consistent training procedure: we iteratively transport observed data to clean samples via a learned velocity field, then enforce consistency by requiring that these generated samples, when passed through $\mathcal{F}$, reproduce the original observation distribution. This circumvents the need for clean training data while extending classical generative modeling to inverse problem settings where only indirect measurements are available.

1.1 Problem setup

We consider a probability distribution of interest $\pi \in \mathcal{P}(\Omega)$, and a forward model $\mathcal{F} : \Omega \to \tilde{\Omega}$ that we allow to be stochastic, i.e., $y = \mathcal{F}(x)$ defines a conditional distribution of y given x on

NeurIPS 2025 Workshop: Frontiers in Probabilistic Inference: Sampling Meets Learning.

$\tilde{\Omega}$. Some representative examples in $\Omega = \mathbb{R}^d$ include the additive white gaussian noise (AWGN) channel $y = x + \sigma\xi$, with $\xi \sim \gamma_d \equiv \mathcal{N}(0, \mathrm{I}_d)$, or tomography, where $y = M\tilde{x} + \sigma\xi$, $\tilde{x}$ is the Fourier transform of x and M is a certain (possibly random) projection operator along frequency rays.

Since $\mathcal{F}$ is a channel that does not introduce additional information about x, we assume that the observation space $\tilde{\Omega}$ can be embedded back into the data space Ω without loss of information. With a slight abuse of notation, we therefore redefine $\mathcal{F}$ as a map $\mathcal{F} : \Omega \to \Omega$. We define the kernel $k_{\mathcal{F}}(y, x)$ associated with $\mathcal{F}$ as the conditional distribution of $y = \mathcal{F}(x)$ given x. This channel pushes forward the data distribution π to an observation distribution $\mu \in \mathcal{P}(\Omega)$, given by $\mu = \mathcal{K}_{\mathcal{F}}\pi$, where $\mathcal{K}_{\mathcal{F}}$ is the integral operator with kernel $k_{\mathcal{F}}$, i.e., $\mu(y) = \int k_{\mathcal{F}}(y, x)\, \mathrm{d}\pi(x)$.

The forward operator $\mathcal{F}$ is often ill-conditioned, non-deterministic (and therefore non-invertible) as a mapping in Ω, thus justifying the need to regularize the inverse problem of recovering x from the observations $y = \mathcal{F}(x)$. However, the situation is different when viewed at the level of probability measures $\mathcal{P}(\Omega)$: as soon as $\mathcal{K}$ is invertible in $\mathcal{P}(\Omega)$, one can hope to recover π from μ by inverting the linear relationship $\mu = \mathcal{K}_{\mathcal{F}}\pi$. To illustrate this point, consider the AWGN channel: while the optimum reconstruction at the level of the samples (in the MSE sense) is given by the posterior mean $\hat{x} = \mathbb{E}[x|\mathcal{F}(x)]$, and generically we have $\mathbb{E}\,\|x - \hat{x}\|^2 > 0$, the associated inverse problem at the level of distributions amounts to a deconvolution, i.e., $\mu = \pi \star \gamma_\sigma$, which is invertible for any σ.

We approach such 'inverse' generative modeling by first assuming that we have access to μ, either directly, or from a dataset of observations $\{y_i\}_i$, $y_i \sim \mu$ that can be fed into a generative modeling that produces an estimate $\hat{\mu}$. We also assume only black-box access to a general (potentially nonlinear) forward model, without requiring its analytical form or gradients.

Related works. In contrast, most existing works impose stronger assumptions on the forward model. For example, Daras et al. [2023], Kawar et al. [2024] train diffusion models with corrupted data under explicit linear forward models and additional full-rank condition. Akyildiz et al. [2025] learns a generative prior by minimizing the sliced-Wasserstein-2 distance between observed data and model outputs. A classical alternative is variational inference, such as the EM algorithm used in Rozet et al. [2024]. On the theoretical side, Li et al. [2024, 2025] study inverse problems over measure spaces, analyzing stability, variational structures, and gradient flows.

2 Stochastic Interpolants for Inverse Generative Modeling

2.1 Standard generative nodel (assuming access to clean data)

Let π be the clean data distribution we wish to sample from and μ be the distribution of the corrupted data that are available to us, both supported on $\mathbb{R}^d$. Following Albergo et al. [2023], a linear stochastic interpolant I_t between π and μ is defined by

$$I_t = \alpha_t x_0 + \beta_t x_1 + \gamma_t z, \quad t \in [0, 1], \tag{1}$$

where (x_0, x_1) is sampled from a joint distribution (or coupling) $\nu(\mathrm{d}x_0, \mathrm{d}x_1)$ that maintains the marginals $\int_{\mathbb{R}^d} \nu(\cdot, \mathrm{d}x_1) = \pi$, $\int_{\mathbb{R}^d} \nu(\mathrm{d}x_0, \cdot) = \mu$, and $z \sim \gamma_d$ is independent Gaussian noise. The schedules $\alpha_t, \beta_t, \gamma_t$ satisfy boundary conditions $\alpha_0 = \beta_1 = 1$, $\alpha_1 = \beta_0 = 0$, and $\gamma_0 = \gamma_1 = 0$. Define the velocity field $b(t, x) \coloneqq \mathbb{E}[\dot{I}_t | I_t = x]$. The solutions to the probability flow ODE

$$\dot{X}_t = b(t, X_t) \tag{2}$$

have the property that $X_{t=1} \sim \mu$ if $X_{t=0} \sim \pi$ (forward direction), and $X_{t=0} \sim \pi$ if $X_{t=1} \sim \mu$ (backward direction). The latter enables clean sample generation from the observation distribution by integrating backward using the drift b. The drift b can be learned efficiently in practice by solving the least-squares regression problem

$$\arg\min_{\hat{b}} \int_0^1 \mathbb{E}[|\hat{b}(t, I_t) - \dot{I}_t|^2]\, \mathrm{d}t, \tag{3}$$

where $\mathbb{E}$ denotes an expectation over the coupling $(x_0, x_1) \sim \nu$ and z.

In the standard generative modeling setting with direct access to clean data samples $x_0 \sim \pi$ and corrupted samples $x_1 \sim \mu$, one may use, for example, the independent coupling $\nu(\mathrm{d}x_0, \mathrm{d}x_1) =$

$\pi(\mathrm{d}x_0)\mu(\mathrm{d}x_1)$ to construct a Monte Carlo approximation of the expectation in the objective (3). *However, in our inverse problem setting, we only observe corrupted data from μ and lack access to clean samples from π. So it is a priori not obvious how to construct the SI* (1). Next we explain how to proceed assuming access to the forward map $\mathcal{F}$.

2.2 Self-consistent generative model (assuming no access to clean data)

Let Φ_b denote the transport map induced by the velocity field b; that is, $\Phi_b(y) = Y_0$ under backward ODE $\dot{Y}_t = b(t, Y_t)$ with terminal condition $Y_1 = y$. The key idea is to view the clean data distribution as the pushforward of the observation distribution μ under the map Φ_b, denoted by $\pi_b := (\Phi_b)_{\#}\mu$, and require that this inferred π_b reproduces the observation distribution under the forward model $\mathcal{F}$: $\mathcal{K}_{\mathcal{F}}\pi_b = \mu$. To implement this, we define a new SI between $\Phi_b(y)$ and $\mathcal{F}(\Phi_b(y))$

$$I_t^b = \alpha_t \Phi_b(y) + \beta_t \mathcal{F}(\Phi_b(y)) + \gamma_t z, \quad t \in [0,1],\ y \sim \mu,\ z \sim \gamma_d,\ y \perp z, \tag{4}$$

where the superscript b emphasizes dependence on the transport map. We then minimize a modified version of the SI loss:

$$\mathcal{L}(\hat{b}) := \int_0^1 \mathbb{E}[|\hat{b}(t, I_t^{\hat{b}}) - \dot{I}_t^{\hat{b}}|^2]\,\mathrm{d}t. \tag{5}$$

The minimizer b^* of the above objective gives us the desired velocity that transports corrupted samples from μ into clean samples from π. To see this, consider $\pi_{b^*} := (\Phi_{b^*})_{\#}\mu$. By definition, b^* transports π_{b^*} to μ under the forward probability ODE (2). On the other hand, $I_t^{b^*}$ can be viewed as a SI between π_{b^*} and $\mathcal{K}_{\mathcal{F}}\pi_{b^*}$. As a result we must have $\mathcal{K}_{\mathcal{F}}\pi_{b^*} = \mu$, which means that $\pi_{b^*} = \pi$.

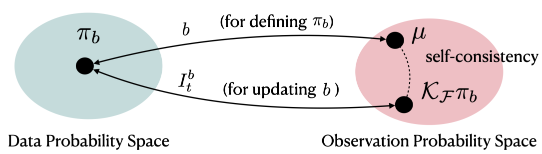

Figure 1: Schematic of the method: the minimizer b^* of the objective (5) satisfies $\mathcal{K}_{\mathcal{F}}\pi_{b_*} = \mu$, which in turns implies that $\pi_{b_*} = \pi$. Note that no samples from π are required—the approach only uses corrupted samples from μ and the map $\mathcal{F}$.

2.3 Practical optimization

The loss function (5) depends on the velocity field b in two ways: first, through the stochastic interpolant I_t^b, which involves the transport map Φ_b applied to the observations; and second, through the regression term itself. Optimizing this objective directly requires differentiating through the transport map, leading to high computational cost and instability during training.

To address this, we adopt an iterative scheme that alternates between two steps: (1) sampling SI using a frozen copy $\tilde{b}$ of the velocity field, and (2) updating the parameters θ of a trainable field b_θ by minimizing the regression loss. Every T_{tr} steps, we update $\tilde{b} \leftarrow b_\theta$. See Algorithm 1 for a detailed description. We remark that in the special case $T_{\text{tr}} = 1$, the algorithm is equivalent to applying stop-gradient to I_t^b, and in some cases we find choosing $T_{\text{tr}} > 1$ leads to better performance.

3 Experimental Case Studies

We present results on imaging tasks with three different forward models, each highlighting a distinct aspect of our approach. Using the CIFAR-10 dataset as the clean data distribution π, we generate one observation y per image with the forward model, resulting in 50,000 observation samples. To optimize the velocity field b in our SI, we use the same U-Net architecture as Dhariwal and Nichol [2021] but with only 64 channels[1], resulting in ∼30 million parameters.

i) Random masking: Following Daras et al. [2023], Rozet et al. [2024], this map generates an observation y by masking each pixel of an image x independently with probability ρ, and adding negligible isotropic Gaussian noise with covariance $\Sigma = 10^{-6}$. As in their setting, we assume access to the mask M for each y and use it to condition our SI. We also pre-process the observations by adding an independent standard Gaussian noise to masked pixels as it improves the final results.

[1] Specifically, we use the implementation here.

Algorithm 1: Self-Consistent Generative Modeling

Input : Observation distribution μ, Forward mapping $\mathcal{F}$, Interpolant schedule (α, β, γ), Drift network b_θ, Number of transport steps T_{tr}, Total number of iterations T

Output : Optimized drift network

$\tilde{b} = b_\theta$ `// Initialize transport map`

for i *in* $1 \ldots T$ **do**

  $y \sim \mu$

  $x = \Phi_{\tilde{b}}(y)$ `// Backward ODE integration to get a data sample`

  $\tilde{y} = \mathcal{F}(x)$ `// Map back to observations`

  $z \sim \mathcal{N}(0, 1);\ t \sim \mathcal{U}(0, 1)$

  $I_t = \alpha_t x + \beta_t \tilde{y} + \gamma_t z$

  $\mathcal{L}(b) = \mathbb{E}\,||b_\theta(t, I_t) - \dot{I}_t||^2$

  $\theta \leftarrow \theta - \epsilon \nabla_\theta \mathcal{L}$ `// ADAM update`

  if $i \,\%\, T_{\text{tr}} == 0$ **then** $\tilde{b} = b_\theta$ `// Update transport map`

return b_θ

(a) $\mathcal{F}$: Random mask ($\rho = 0.5$).

(b) $\mathcal{F}$: Gaussian blur + noise ($\sigma_R = 1, \sigma_n = 0.1$).

(c) $\mathcal{F}$: Motion blur (kernel size=7)

Figure 2: Restored samples for different forward maps.

Our SI, based on a finite dataset, is not itself a generative model. To compare with generative models from prior work, we use the trained SI to restore the observations; that is, we transport all observations y to the data space via $\Phi_b(y)$, and use these samples to train a generative diffusion model. We use the same architecture as above, but with 96 channels. Table 1 shows the FID scores for observations with two different masking probabilities. Our method vastly outperforms Daras et al. [2023]. It is comparable with Rozet et al. [2024], but more computationally efficient and required a combined 120 GPU hours on A100 compared to their 512 GPU hours.

Method	Corruption level	FID
Ambient Diffusion	0.20	11.70
	0.40	18.85
EM Posterior	0.25	5.88
	0.50	6.76
Ours (generated)	0.25	**5.38**
	0.50	**6.74**
Baseline	0.00	5.16

Table 1: FIDs for random masking. To account for architectures, 'baseline' is FID for our model on clean CIFAR-10 data.

ii) Gaussian blurring with noise: The forward map consists of blurring with a Gaussian kernel with $\sigma = 1$, followed by adding Gaussian noise with $\sigma = 0.10$. This demonstrates that unlike previous works, e.g., Daras et al. [2023], our approach can handle non-negligible noise.

iii) Motion blurring: The previous two examples involve linear forward maps. We now consider a nonlinear one: motion blur. Figure 2c shows restored samples for observations with a 7-pixel motion kernel and small Gaussian noise ($\Sigma = 10^{-6}$). The blur direction is randomly assigned per image and assumed known for conditioning the SI. While Daras et al. [2023], Rozet et al. [2024] are limited to linear operators, our method handles nonlinear maps with only black-box access.

References

O Deniz Akyildiz, Mark Girolami, Andrew M Stuart, and Arnaud Vadeboncoeur. Efficient prior calibration from indirect data. *SIAM Journal on Scientific Computing*, 47(4):C932–C958, 2025.

Michael Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *ICLR 2023 Conference*, 2023.

Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. Ambient diffusion: Learning clean distributions from corrupted data. *Advances in Neural Information Processing Systems*, 36:288–313, 2023.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Bahjat Kawar, Noam Elata, Tomer Michaeli, and Michael Elad. GSURE-based diffusion model training with corrupted data. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Qin Li, Maria Oprea, Li Wang, and Yunan Yang. Stochastic inverse problem: stability, regularization and wasserstein gradient flow. *arXiv preprint arXiv:2410.00229*, 2024.

Qin Li, Maria Oprea, Li Wang, and Yunan Yang. Inverse problems over probability measure space. *arXiv preprint arXiv:2504.18999*, 2025.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. *Advances in Neural Information Processing Systems*, 37:87647–87682, 2024.

Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.