Stochastic Interpolants

Based on "Stochastic Interpolations: A Unifying Framework for Flows and Diffusions" (2025) by Albergo, Boffi, Vanden-Eijnden

Garrett Mulcahy
gmulcahy@uw.edu
https://garrett-mulcahy.github.io/

February 17 2026

UNIVERSITY of WASHINGTON

Contents

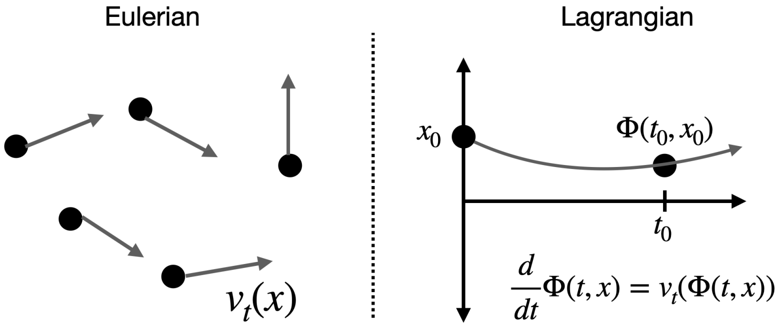
Review from Last Week

Continuity Equation

Continuity Equation

Let $(v_t, t \geq 0)$ be a time dependent velocity field, then

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0.$$



Fokker Planck

Fokker Planck PDE

Let $(b_t, t \geq 0)$ be smoothly varying vector valued function, and recall $\Delta = \sum_{j=1}^{d} \partial_{x_j}^2$, then

$$\partial_t \rho_t + \nabla \cdot (b_t \rho_t) = \varepsilon \Delta \rho_t.$$

Observe that this writes as

$$\partial_t \rho_t + \nabla \cdot \left(\left(b_t^{FP} - \varepsilon \nabla \log \rho_t \right) \rho_t \right) = 0.$$

It holds that $X_t \sim \rho_t$, where

$$dX_t = b_t^{FP}(X_t)dt + \sqrt{2\varepsilon} dB_t, \quad X_0 \sim \rho_0$$

Stochastic Interpolation

Stochastic Interpolation [ABVE25]

Let ρ_0, ρ_1 be two probability densities, a stochastic interpolant between ρ_0 and ρ_1 is the stochastic process defined as

$$x_t := I(t, x_0, x_1) + \gamma(t)z, \quad t \in [0, 1],$$

where

- $I(0, x_0, x_1) = x_0$, $I(1, x_0, x_1) = x_1$, $\gamma(0) = \gamma(1) = 0$,
- $(x_0, x_1) \sim \nu$ that is coupling of ρ_0, ρ_1
- $z \sim N(0, \mathrm{Id})$ is independent of (x_0, x_1)

Stochastic Interpolation

Specific choices of $x_t = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z$

Stochastic Interpolant		$\alpha(t)$	$\beta(t)$	$\gamma(t)$
Arbitrary ρ_0 (two-sided)	linear	$1-t$	t	$\sqrt{at(1-t)}$
	trig	$\cos\frac{\pi}{2}t$	$\sin\frac{\pi}{2}t$	$\sqrt{at(1-t)}$
	enc-dec	$\cos^2(\pi t)1_{[0,\frac{1}{2})}(t)$	$\cos^2(\pi t)1_{(\frac{1}{2},1]}(t)$	$\sin^2(\pi t)$
Gaussian ρ_0 (one-sided)	linear	$1-t$	t	0
	trig	$\cos\frac{\pi}{2}t$	$\sin\frac{\pi}{2}t$	0
	SBDM (VP)	$\sqrt{1-t^2}$	t	0
Mirror		0	1	$\sqrt{at(1-t)}$

Figure: [ABVE25, Table 1], note “one-sided” setting includes diffusion models

Stochastic Interpolation

Given $x_t = I(t, x_0, x_1) + \gamma(t)z$, let $x_t \sim \rho_t$

Transport Equation

The PDE $\partial_t \rho_t + \nabla \cdot (b_t \rho_t) = 0$ is satisfied with

$$b(t, x) = \mathrm{E}[\dot{x}_t | x_t = x] = \mathrm{E}[\partial_t I(t, x_0, x_1) + \dot{\gamma}(t)z | x_t = x]$$

Fokker Planck Equation

The PDE $\partial_t \rho_t + \nabla \cdot (b_t^{FP} \rho_t) = \varepsilon \Delta \rho_t$ is satisfied with

$$s(t, x) = \nabla \log \rho(t, x) = -\gamma^{-1}(t)\mathrm{E}[z | x_t = x]$$
$$b^{FP} = b + \varepsilon s$$

Choice of Coupling

What's the point of choosing $(X_0, X_1) \sim \nu$? Another **design choice**

- Diffusion models [SSDK+20], one marginal is noise and assume independent
- [AGB+24] analyzes theoretical and practical advantages of ν selection
- Another meaningful coupling is Schrödinger bridge [SDBCD23, DBKMD24]

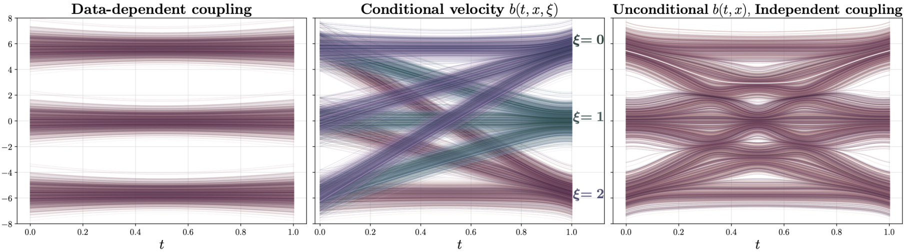


Figure: [AGB+24, Figure 3], Gaussian mixture with 3 modes, different couplings of endpoints

Deterministic vs Stochastic Trajectories

Two Approaches to Generative Modeling

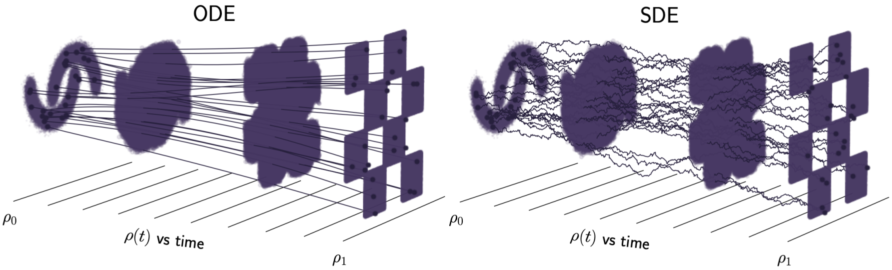


Figure: ODE vs SDE Generative Modeling [ABVE25, Figure 1]

An Algorithmic Question: Which one is "better"?

Generative Modeling (ODE)

Deterministic transport (ODE), estimate $b(t,x)$ such that

$$\partial_t \rho_t + \nabla \cdot (b_t \rho_t) = 0 \tag{1}$$

[ABVE25, Lemma 21], Endpoint Error

Consider $\rho(t,x), \hat{\rho}(t,x)$ satisfying (1) with $b(t,x), \hat{b}(t,x)$, respectively. Then

$$\mathrm{KL}(\rho(1)|\hat{\rho}(1)) = \int_0^1 \int_{\mathbb{R}^d} \left\langle \nabla \log \hat{\rho}(t,x) - \nabla \log \rho(t,x), \hat{b}(t,x) - b(t,x) \right\rangle \rho(t,x) dx dt$$

- Key ideas in proof are: (1) interpolation in time, (2) integration by parts :)
- **Problem:** error not controlled entirely by $(\hat{b} - b)$

Generative Modeling (ODE)

As $\rho(0,\cdot) = \hat{\rho}(0,\cdot)$,

$$\mathrm{KL}(\rho(1)|\hat{\rho(1)}) = \mathrm{KL}(\rho(1)|\hat{\rho(1)}) - \mathrm{KL}(\rho(0)|\hat{\rho(0)}) = \int_0^1 \frac{d}{dt}\mathrm{KL}(\rho(t)|\hat{\rho}(t))dt.$$

Recall that $\mathrm{KL}(p|q) = \mathrm{E}_p[\log(dp/dq)]$, so

$$\begin{aligned}\frac{d}{dt}\mathrm{KL}(\rho(t)|\hat{\rho}(t)) &= \frac{d}{dt}\int_{\mathbb{R}^d} \log\left(\frac{\rho(t,x)}{\hat{\rho}(t,x)}\right)\rho(t,x)dx \\ &= \int_{\mathbb{R}^d} \partial_t\left(\log\left(\frac{\rho(t,x)}{\hat{\rho}(t,x)}\right)\right)\rho(t,x) + \left(\log\left(\frac{\rho(t,x)}{\hat{\rho}(t,x)}\right)\right)\partial_t\rho(t,x)dx\end{aligned}$$

Recall that $\partial_t\rho(t) = -\nabla\cdot(\rho(t)b(t))$, so integration by parts gives

$$\int_{\mathbb{R}^d}\left(\log\left(\frac{\rho(t,x)}{\hat{\rho}(t,x)}\right)\right)\partial_t\rho(t,x)dx = \int_{\mathbb{R}^d}\langle\nabla\log\rho(t) - \nabla\log\hat{\rho}(t), b(t)\rangle\rho(t,x)dx$$

Generative Modeling (ODE)

For the other term,

$$
\begin{aligned}
&\int_0^1 \int_{\mathbb{R}^d} \big(\partial_t \log \rho(t) - \partial_t \log \hat{\rho}(t)\big)\, \rho(t,x) dx dt \\
&= \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\partial_t \rho(t)}{\rho(t)} - \frac{\partial_t \hat{\rho}(t)}{\hat{\rho}(t)} \right) \rho(t,x) dx dt \\
&= \int_0^1 \int_{\mathbb{R}^d} \left(-\frac{\nabla \cdot (\rho(t) b(t))}{\rho(t)} + \frac{\nabla \cdot (\hat{\rho}(t)\hat{b}(t))}{\hat{\rho}(t)} \right) \rho(t,x) dx dt \\
&= \int_0^1 \left(-\int_{\mathbb{R}^d} \left(\nabla \cdot (b(t) - \hat{b}(t)) \right) \rho(t,x) dx \right) dt \\
&+ \int_0^1 \int_{\mathbb{R}^d} \left(-\langle \nabla \log \rho(t), b(t) \rangle + \langle \nabla \log \hat{\rho}(t), \hat{b}(t) \rangle \right) \rho(t,x) dx dt.
\end{aligned}
$$

Generative Modeling (ODE)

Using integration by parts,

$$-\int_{\mathbb{R}^d} \left(\nabla \cdot (b(t) - \hat{b}(t))\right) \rho(t,x)dx = \int_{\mathbb{R}^d} \langle b(t) - \hat{b}(t), \nabla \log \rho(t) \rangle \rho(t,x)dx,$$

so previous slide becomes

$$\begin{aligned}
&\int_0^1 \int_{\mathbb{R}^d} \left(\partial_t \log \rho(t) - \partial_t \log \hat{\rho}(t)\right) \rho(t,x)dxdt \\
&= \int_0^1 \int_{\mathbb{R}^d} \langle b(t) - \hat{b}(t), \nabla \log \rho(t) \rangle \rho(t,x)dxdt \\
&+ \int_0^1 \int_{\mathbb{R}^d} \left(-\langle \nabla \log \rho(t), b(t) \rangle + \langle \nabla \log \hat{\rho}(t), \hat{b}(t) \rangle\right) \rho(t,x)dxdt \\
&= \int_0^1 \langle \hat{b}(t), \nabla \log \hat{\rho}(t) - \nabla \log \rho(t) \rangle \rho(t,x)dxdt.
\end{aligned}$$

Combing terms gives final solution!

Generative Modeling (SDE)

Fokker-Planck, estimate $b(t, x)$ such that

$$\partial_t \rho_t + \nabla \cdot (b_t^{FP} \rho_t) = \varepsilon \Delta \rho_t \tag{2}$$

[ABVE25, Lemma 22], Endpoint Error

Consider $\rho(t, x), \hat{\rho}(t, x)$ satisfying (2) with $b^{FP}(t, x), \hat{b}^{FP}(t, x)$, respectively. Then

$$\mathrm{KL}(\rho(1)|\hat{\rho}(1)) \leq \frac{1}{4\varepsilon} \int_0^1 \int_{\mathbb{R}^d} \left| b^{FP}(t, x) - \hat{b}^{FP}(t, x) \right|^2 \rho(t, x) dx dt$$

Generative Modeling (SDE)

Recall that

$$\partial_t \rho_t + \nabla \cdot (b_t^{FP} \rho_t) = \varepsilon \Delta \rho_t \Leftrightarrow \partial_t \rho_t + \nabla \cdot \left(\left(b_t^{FP} - \varepsilon \nabla \log \rho_t \right) \rho_t \right) = 0$$

From previous calculation,

$$\begin{aligned}
&\mathrm{KL}(\rho(1)|\hat{\rho}(1)) \\
&= \int_0^1 \int_{\mathbb{R}^d} \left\langle \nabla \log \hat{\rho}_t - \nabla \log \rho_t, \hat{b}_t^{FP} - \varepsilon \nabla \log \hat{\rho}_t - b_t^{FP} + \varepsilon \nabla \log \rho_t \right\rangle \rho_t(x) dx dt \\
&= \int_0^1 \int_{\mathbb{R}^d} \left(\langle \nabla \log \hat{\rho}_t - \nabla \log \rho_t, \hat{b}_t^{FP} - b_t^{FP} \rangle - \varepsilon \|\nabla \log \hat{\rho}_t - \nabla \log \rho_t\|^2 \right) \rho_t(x) dx dt \\
&\leq \frac{1}{2\eta^2} \int_0^1 \left\| \hat{b}_t^{FP} - b_t^{FP} \right\|^2 \rho_t(x) dx dt + \left(\frac{\eta^2}{2} - \varepsilon \right) \int_0^1 \|\nabla \log \hat{\rho}_t - \nabla \log \rho_t\|^2 \rho_t(x) dx dt,
\end{aligned}$$

using Young's inequality. Then set $\eta = \sqrt{2\varepsilon}$.

Generating Modeling, [ABVE25, Remark 25]

[ABVE25, Theorem 23], Endpoint Error

In the setting as above, let $\hat{b}$ be estimate for b, then

$$\mathrm{KL}(\rho(1)|\hat{\rho}(1)) \leq \frac{1}{2\varepsilon}\left(\text{Error in } \hat{b}\right) + \frac{\varepsilon}{2}\left(\text{Error in } \hat{s}\right)$$

Suppose we have estimates $\hat{b}, \hat{s}$

- Deterministic generative model

$$\frac{d}{dt}\hat{X}_t = \hat{b}(t, \hat{X}_t)$$

- Stochastic generative model

$$d\hat{X}_t = \left(\hat{b}(t, \hat{X}_t) + \varepsilon\hat{s}(t, \hat{X}_t)\right)dt + \sqrt{2\varepsilon}dW_t$$

- The optimal value of ε is

$$\varepsilon^* = \left(\frac{(\text{Error in } \hat{b})}{(\text{Error in } \hat{s})}\right)^{1/2}$$

Generative Modeling

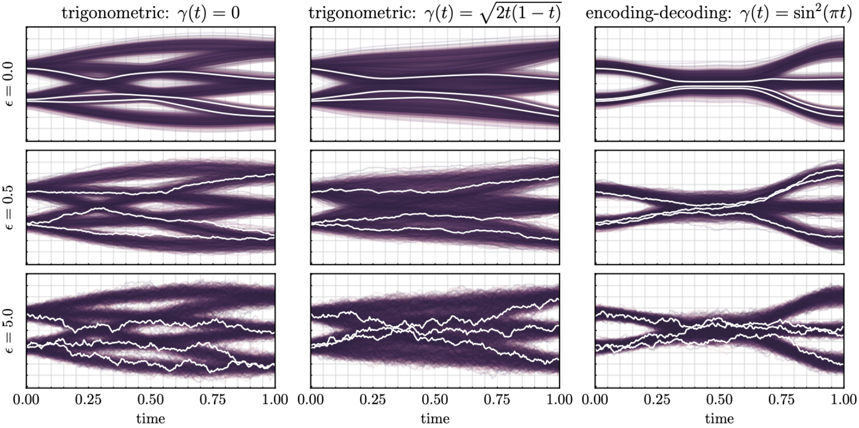


Figure: [ABVE25, Figure 5], Role of ε on sample trajectories

[ABVE25] Michael Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

[AGB+24] Michael S Albergo, Mark Goldstein, Nicholas M Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. In *Proceedings of the 41st International Conference on Machine Learning*, pages 921–937, 2024.

[DBKMD24] Valentin De Bortoli, Iryna Korshunova, Andriy Mnih, and Arnaud Doucet. Schrodinger bridge flow for unpaired data translation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 103384–103441. Curran Associates, Inc., 2024.

[SDBCD23] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 36:62183–62223, 2023.

[SSDK+20] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020. ArXiv version 2011.13456.