Rectified Flow and Stochastic Interpolation

Siqi Yao

SCHOOL OF DATA SCIENCE

July 13, 2025

Table of Contents

Table of Contents

Transport Mapping Problem

- Generative modelling and data transfer problems can be formulated as **learning transport mapping:**
  Given empirical observations of two distributions π_0, π_1 on $\mathbb{R}^d$, find a transport map $T : \mathbb{R}^d \to \mathbb{R}^d$, which, in the infinite data limit, gives $Z_1 := T(Z_0) \sim \pi_1$ when $Z_0 \sim \pi_0$.
- Examples:
  - Latent variable $\xrightarrow{\text{implicit map}}$ samples: GAN, VAE.
  - Prior distribution $\xrightarrow{\text{explicit flow}}$ samples: Flow and diffusion models.
  - We will focus on the **explicit flow-based** approach.

Simplest Path: Straight Path

- Suppose we have some $X_0 \sim \pi_0$ (prior), $X_1 \sim \pi_1$ (samples) and seek to transform X_0 to X_1.
- A key issue with neural-ODEs is their complex trajectories.
- However, since the paths can be **arbitrary**, a natural idea is to simply connect randomly sampled pairs (X_0, X_1) with **straight lines**, as demonstrated in the Figure 1 and 2.

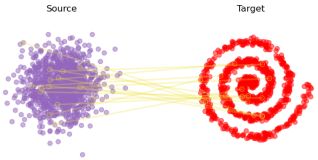

Figure 1: 2D View.

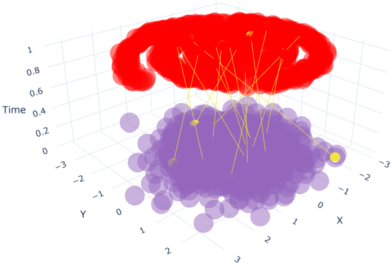

Figure 2: 3D View.

Problems with Straight Paths

- Despite its simplicity, this approach encounters several fundamental challenges:
  - **Crossings:** The interpolation paths X_t can intersect, resulting in non-unique solutions, as shown in Figure 3.
  - **Non-causality:** These paths are constructed using both starting and ending points, but in reality, the true endpoints (samples) are unknown during inference.

Figure 3: Linear Interpolation. Source: [Liu et al., 2022]

Intuition of Rectified Flow

- In rectified flow, data points evolve according to an ODE:

$$\mathrm{d}Z_t = v(Z_t, t)\,\mathrm{d}t, \qquad t \in [0, 1] \tag{1}$$

  where velocity field v is a **parameterized network**, ensuring that the ODE trajectory closely follows the **imagined** straight path.
- ODEs inherently prevent trajectory crossings: by the **Picard–Lindelöf theorem**, each starting point generates a unique path, provided v is Lipschitz continuous.
- The resulting flow can be seen as an "average" of many particles traveling along these imagined straight lines, blending their crossings into a smooth, non-crossing transformation.

Table of Contents

Model

- Given X_0 and X_1, we have the **non-causal** linear interpolation:

$$X_t = tX_1 + (1-t)X_0, \qquad t \in [0,1] \tag{2}$$

- The **rectified flow** induced from empirical observations (X_0, X_1) is an ODE:

$$\mathrm{d}Z_t = v(Z_t, t)\,\mathrm{d}t, \qquad t \in [0,1] \tag{3}$$

  which converts Z_0 from π_0 to Z_1 following π_1.
- $v : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$ is trained to drive the flow to follow the direction $(X_1 - X_0)$ of the linear path pointing from X_0 to X_1, by solving a least squares regression problem:

$$\min_v \int_0^1 \mathbb{E}\left[\left\|(X_1 - X_0) - v(X_t, t)\right\|^2\right] \mathrm{d}t, \qquad X_t = tX_1 + (1-t)X_0 \tag{4}$$

  which is estimated by randomly samplings pairs of (X_0, X_1) and t, and optimized using SGD.

Algorithm

- **Procedure:** $\boldsymbol{Z} = \text{RectFlow}((X_0, X_1))$:
  - **Inputs:** Velocity model $v_\theta : \mathbb{R}^d \to \mathbb{R}^d$ with parameter θ.
  - **Training:** $\hat{\theta} = \arg\min_\theta \mathbb{E}\left[\left\| X_1 - X_0 - v\big(tX_1 + (1-t)X_0, t\big)\right\|^2\right]$, where $t \sim \text{Uniform}([0,1])$, $(X_0, X_1) \sim \pi_0 \times \pi_1$.
  - **Sampling:** Draw (Z_0, Z_1) following $\mathrm{d}Z_t = v_{\hat{\theta}}(Z_t, t)\mathrm{d}t$ starting from $Z_0 \sim \pi_0$ (or $Z_1 \sim \pi_1$).
  - **Return:** $\boldsymbol{Z} = \{Z_t : t \in [0,1]\}$.
- **Reflow (optional):** $\boldsymbol{Z}^{k+1} = \text{RectFlow}((Z_0^k, Z_1^k))$, starting from $(Z_0^0, Z_1^0) = (X_0, X_1)$, where $(X_0, X_1) \sim \pi_0 \times \pi_1$.
- **Distill (optional):** Learn a neural network $\hat{T}$ to distill the k-rectified flow, such that $Z_1^k \approx \hat{T}(Z_0^k)$.

Table of Contents

Flows Avoid Crossing

- A well-defined ODE yields unique solutions: **the trajectories cannot cross at any time** $t \in [0, 1]$.
- This uniqueness prevents multiple paths from passing through the same point at the same time in different directions (Figure 4).
- Think of linear interpolation X_t as "building roads" between π_0 and π_1, while rectified flow finds an **averaged** path without crossing (Figure 5).

Figure 4: Crossings.

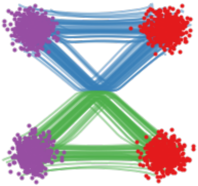

Figure 5: No Crossings.

Minimizer of Objective

Important note: in all context, a **coupling** (X_0, X_1) where $X_0 \sim \pi_0, X_1 \sim \pi_1$, stands for a **joint distribution** $\gamma(X_0, X_1)$, whose marginal distribution are π_0, π_1 respectively.

Theorem 1

For a given input coupling (X_0, X_1), the exact minimum of (4) is achieved by

$$v^X(x,t) = \mathbb{E}[X_1 - X_0 \mid X_t = x] = \int (X_1 - X_0)\, P_{(X_0,X_1)|X_t=x}(dX_0, dX_1) \tag{5}$$

which is the expectation of the line directions $X_1 - X_0$ that pass through x at time t.

This is based on **the following result:** Let $\hat{X} = g(Y)$ be an estimator of the random variable X. The MSE of this estimator is defined as

$$E[(X - \hat{X})^2] = E\left[(X - g(Y))^2\right]. \tag{6}$$

The MMSE (Minimum Mean Squared Error) estimator of X,

$$\hat{X}_M = E[X \mid Y], \tag{7}$$

has the lowest MSE among all possible estimators.

Nonlinear Extension of Rectified Flow

- We consider a nonlinear extension of rectified flow, where the linear interpolation X_t is replaced by any time-differentiable curve connecting X_0 and X_1.
- Let $\boldsymbol{X} = \{X_t : t \in [0,1]\}$ be any time-differentiable random process connecting X_0 and X_1, with time derivative $\frac{\mathrm{d}X_t}{\mathrm{d}t} := \dot{X}_t$. The nonlinear rectified flow induced from $\boldsymbol{X}$ is defined as:

$$\mathrm{d}Z_t = v^{\boldsymbol{X}}(Z_t, t)\,\mathrm{d}t, \text{ with } Z_0 = X_0, \qquad v^{\boldsymbol{X}}(z,t) = \mathbb{E}[\dot{X}_t \mid X_t = z]. \tag{8}$$

- $v^{\boldsymbol{X}}$ can be considered as the **optimization goal**, and we find it by solving:

$$\min_v \int_0^1 \mathbb{E}\left[w_t \left\|\dot{X}_t - v(X_t,t)\right\|^2\right] \mathrm{d}t, \tag{9}$$

where $w_t : (0,1) \to (0,+\infty)$ is a positive weighting sequence ($w_t = 1$ by default). When $X_t = tX_1 + (1-t)X_0$ and $w_t = 1$, (9) becomes (4).

Marginal Preserving Property

The marginal preserving property that $\text{Law}(Z_t) = \text{Law}(X_t) \quad \forall t$ is a general property of the nonlinear rectified flow in (8). **This is why we need to find** $v^{\boldsymbol{X}}$.

Definition 1

For a path-wise continuously differentiable random process $\boldsymbol{X} = \{X_t : t \in [0,1]\}$, its *expected velocity* $v^{\boldsymbol{X}}$ is defined as

$$v^{\boldsymbol{X}}(x,t) = \mathbb{E}[\dot{X}_t \mid X_t = x], \qquad \forall x \in \text{supp}(X_t). \tag{10}$$

For $x \notin \text{supp}(X_t)$, the conditional expectation is not defined and we set $v^{\boldsymbol{X}}$ arbitrarily, say $v^{\boldsymbol{X}}(x,t) = 0$.

Definition 2

We call that $\boldsymbol{X}$ is *rectifiable* if $v^{\boldsymbol{X}}$ is locally bounded and the solution of the integral equation below exists and is unique:

$$Z_t = Z_0 + \int_0^t v^{\boldsymbol{X}}(Z_t, t)\mathrm{d}t, \qquad \forall t \in [0,1], \qquad Z_0 = X_0. \tag{11}$$

In this case, $\boldsymbol{Z} = \{Z_t : t \in [0,1]\}$ is called the *rectified flow induced from* $\boldsymbol{X}$.

Marginal Preserving Property (cont'd)

Theorem 2

Assume $\boldsymbol{X}$ is rectifiable and $\boldsymbol{Z}$ is its rectified flow. Then $\mathrm{Law}(Z_t) = \mathrm{Law}(X_t) \quad \forall t \in [0,1]$.

Proof.
For any compactly supported continuously differentiable test function $h : \mathbb{R}^d \to \mathbb{R}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[h(X_t)] = \mathbb{E}[\nabla h(X_t)^\top \dot{X}_t] \tag{12}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\nabla h(X_t)^\top \dot{X}_t \mid X_t\right]\right] \tag{13}$$

$$= \mathbb{E}\left[\nabla h(X_t)^\top \mathbb{E}\left[\dot{X}_t \mid X_t\right]\right] \tag{14}$$

$$= \mathbb{E}[\nabla h(X_t)^\top v^{\boldsymbol{X}}(X_t, t)] \tag{15}$$

where in (12) we used chain rule and in (15) we used the definition of $v^{\boldsymbol{X}}(X_t, t)$.
This is equivalent to that $\pi_t := \mathrm{Law}(X_t)$ solves the continuity equation (FP equation) with drift $v_t^{\boldsymbol{X}} := v^{\boldsymbol{X}}(\cdot, t)$:

$$\dot{\pi}_t + \nabla \cdot (v_t^{\boldsymbol{X}} \pi_t) = 0. \tag{16}$$

Marginal Preserving Property (cont'd)

To see this, we can multiply (16) with h and integrate both sides:

$$0 = \int h\big(\dot{\pi}_t + \nabla \cdot (v_t^{\boldsymbol{X}} \pi_t)\big) \mathrm{d}x \tag{17}$$

$$= \int (h\,\dot{\pi}_t - \nabla h^\top v_t^{\boldsymbol{X}} \pi_t) \mathrm{d}x \tag{18}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[h(X_t)] - \mathbb{E}[\nabla h(X_t)^\top v^{\boldsymbol{X}}(X_t, t)], \tag{19}$$

where we use integration by parts that $\int h\,\nabla \cdot (v_t^{\boldsymbol{X}} \pi_t) \mathrm{d}x = -\int \nabla h^\top (v_t^{\boldsymbol{X}} \pi_t) \mathrm{d}x$. $\square$
Because Z_t is driven by the same velocity field $v^{\boldsymbol{X}}$, its marginal law $\mathrm{Law}(Z_t)$ **solves the same equation with the same initial condition** $(Z_0 = X_0)$**.**
Hence, the equivalence of $\mathrm{Law}(Z_t)$ and $\mathrm{Law}(X_t)$ follows if the solution of (16) is unique, which is equivalent to the uniqueness of the solution of $\mathrm{d}Z_t = v^{\boldsymbol{X}}(Z_t, t)\,\mathrm{d}t$.
However, the **joint distributions** of the whole trajectory of Z_t and that of X_t are different in general.

Reducing Convex Transport Costs

- This property only holds for **linear** rectified flow.
- Monge's Optimal Transport problem ([Monge, 1781]):

$$\min_T \mathbb{E}\big[c(Z_1 - Z_0)\big] \qquad \text{s.t.} \qquad Z_1 = T(Z_0),\ \mathrm{Law}(Z_0) = \pi_0,\ \mathrm{Law}(Z_1) = \pi_1 \tag{20}$$

where $c : \mathbb{R}^d \to \mathbb{R}$ is a cost function, e.g., $c(x) = \frac{1}{2}\|x\|^2$.

Definition 3

A coupling (X_0, X_1) is called *rectifiable* if its linear interpolation process $\boldsymbol{X} = \{tX_1 + (1-t)X_0 : t \in [0,1]\}$ is rectifiable.
In this case, the $\boldsymbol{Z} = \{Z_t : t \in [0,1]\}$ in (11) is called the *rectified flow* of coupling (X_0, X_1), denoted as $\boldsymbol{Z} = \mathrm{RectFlow}((X_0, X_1))$, and (Z_0, Z_1) is called the *rectified coupling* of (X_0, X_1), denoted as $(Z_0, Z_1) = \mathrm{RectFlow}((X_0, X_1))$.

Theorem 3

Assume (X_0, X_1) is rectifiable and $(Z_0, Z_1) = \mathrm{RectFlow}((X_0, X_1))$. Then for any convex function $c : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathbb{E}[c(Z_1 - Z_0)] \le \mathbb{E}[c(X_1 - X_0)]. \tag{21}$$

Reducing Convex Transport Costs (cont'd)

Proof.

$$\begin{aligned}
\mathbb{E}\big[c(Z_1 - Z_0)\big] &= \mathbb{E}\left[c\left(\int_0^1 v^{\boldsymbol{X}}(Z_t, t)\,\mathrm{d}t\right)\right] &(22)\\
&\le \mathbb{E}\left[\int_0^1 c\left(v^{\boldsymbol{X}}(Z_t, t)\right)\mathrm{d}t\right] &(23)\\
&= \mathbb{E}\left[\int_0^1 c\left(v^{\boldsymbol{X}}(X_t, t)\right)\mathrm{d}t\right] &(24)\\
&= \mathbb{E}\left[\int_0^1 c\left(\mathbb{E}[X_1 - X_0 \mid X_t]\right)\mathrm{d}t\right] &(25)\\
&\le \mathbb{E}\left[\int_0^1 \mathbb{E}\left[c(X_1 - X_0) \mid X_t\right]\mathrm{d}t\right] &(26)\\
&= \int_0^1 \mathbb{E}\left[c(X_1 - X_0)\right]\mathrm{d}t &(27)\\
&= \mathbb{E}\big[c(X_1 - X_0)\big]. &(28)
\end{aligned}$$

where the two inequalities follow from Jensen's inequality. □

Straightening Effect of Reflow

- This property also holds for **linear** rectified flow only.
- $\boldsymbol{Z} = \text{RectFlow}((X_0, X_1))$ denotes the rectified flow induced from (X_0, X_1). **Applying the algorithm recursively** yields a sequence of rectified flows $\boldsymbol{Z}^{k+1} = \text{RectFlow}\big((Z_0^k, Z_1^k)\big)$ with $(Z_0^0, Z_1^0) = (X_0, X_1)$.
- The reflow procedure not only decreases transport cost, but can also **straighten the paths of rectified flows** (Figure 6).
- This is a nice property since perfectly straight paths can be simulated exactly with a single Euler step.

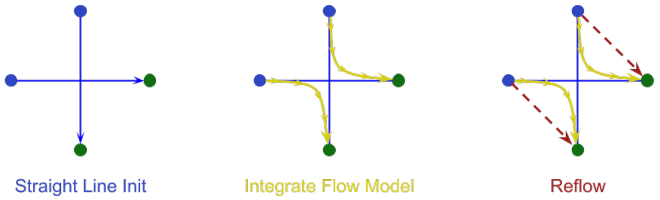

Figure 6: Straightening Effect of Reflow. Source: [Hawley, 2024]

Straightening Effect of Reflow (cont'd)

- The straightness of any smooth process $\boldsymbol{Z} = \{Z_t\}$ can be measured by

$$S(\boldsymbol{Z}) = \int_0^1 \mathbb{E}\left[\left\|(Z_1 - Z_0) - \dot{Z}_t\right\|^2\right] \mathrm{d}t. \tag{29}$$

  $S(\boldsymbol{Z}) = 0$ implies exact straightness, since now $v(Z_t, t) = \dot{Z}_t = Z_1 - Z_0$, meaning that $Z_t = tZ_1 + (1-t)Z_0, \qquad \forall t.$ A flow with a small $S(\boldsymbol{Z})$ has nearly straight paths.
- It is not recommended to apply too many reflow steps as it may accumulate estimation error on $v^{\boldsymbol{X}}$.

Straightening Effect of Reflow (cont'd)

A coupling (X_0, X_1) is said to be straight (or fully rectified) if it is a fixed point of the $\mathrm{RectFlow}(\cdot)$ mapping, i.e., $(X_0, X_1) = \mathrm{RectFlow}\big((X_0, X_1)\big)$. It is desirable to obtain a straight coupling because its rectified flow is straight.

Theorem 4

Assume (X_0, X_1) is rectifiable. Let $X_t = tX_1 + (1-t)X_0$ and $\boldsymbol{Z} = \mathrm{RectFlow}((X_0, X_1))$. Then (X_0, X_1) is a straight coupling iff the following equivalent statements hold.

1. There exists a strictly convex function $c : \mathbb{R}^d \to \mathbb{R}$, such that $\mathbb{E}[c(Z_1 - Z_0)] = \mathbb{E}[c(X_1 - X_0)]$.
2. (X_0, X_1) is a fixed point of $\mathrm{RectFlow}(\cdot)$, i.e., $(X_0, X_1) = (Z_0, Z_1)$. This is the definition of being straight.
3. The rectified flow coincides with the linear interpolation process: $\boldsymbol{X} = \boldsymbol{Z}$.
4. The paths of the linear interpolation $\boldsymbol{X}$ do not intersect:

$$V((X_0, X_1)) := \int_0^1 \mathbb{E}\left[\|X_1 - X_0 - \mathbb{E}[X_1 - X_0 \mid X_t]\|^2\right] \mathrm{d}t = 0, \tag{30}$$

where $V((X_0, X_1)) = 0$ indicates that $X_1 - X_0 = \mathbb{E}[X_1 - X_0 \mid X_t]$ almost surely when $t \sim \mathrm{Uniform}([0, 1])$, meaning that the direction $X_1 - X_0$ for lines passing through each X_t is unique, and hence no linear interpolation paths intersect.

Straightening Effect of Reflow (cont'd)

Proof.

- $3 \to 2 \to 1$: obvious.
- $1 \to 4$: if $\mathbb{E}[c(Z_1 - Z_0)] = \mathbb{E}[c(X_1 - X_0)]$, the two applications of Jensen's inequality in the proof of Theorem 3 are tight. Since c is strictly convex, for any $x \neq y$ and $\lambda \in (0, 1)$ we have

$$c(\lambda x + (1-\lambda)y) < \lambda c(x) + (1-\lambda)c(y) \tag{31}$$

  So the second Jensen's inequality in the proof implies that $X_1 - X_0 = \mathbb{E}[X_1 - X_0 \mid X_t]$ almost surely w.r.t. X and $t \sim \text{Uniform}([0, 1])$, which implies that $V(\boldsymbol{X}) = 0$.
- $4 \to 3$: if $V(\boldsymbol{X}) = 0$, we have $\int_0^s (X_1 - X_0)\mathrm{d}t = \int_0^s \mathbb{E}[X_1 - X_0 \mid X_t]\mathrm{d}t = \int_0^s v^{\boldsymbol{X}}(X_t, t)\mathrm{d}t$ for $s \in (0, 1]$. Hence

$$X_t = X_0 + \int_0^t (X_1 - X_0)\mathrm{d}t = X_0 + \int_0^t v^{\boldsymbol{X}}(X_t, t)\mathrm{d}t. \tag{32}$$

  Because $\boldsymbol{Z}$ satisfies the same equation (11), we have $\boldsymbol{X} = \boldsymbol{Z}$ by the uniqueness of the solution. $\square$

Straightening Effect of Reflow (cont'd)

Theorem 5

Let $\boldsymbol{Z}^k$ be the k-th rectified flow of (X_0, X_1), that is, $\boldsymbol{Z}^{k+1} = \mathrm{RectFlow}((Z_0^k, Z_1^k))$ and $(Z_0^0, Z_1^0) = (X_0, X_1)$. Assume each (Z_0^k, Z_1^k) is rectifiable for $k = 0, \ldots, K$.
Then

$$\sum_{k=0}^{K} \left[S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k)) \right] \leq \mathbb{E}\left[\|X_1 - X_0\|^2 \right]. \tag{33}$$

So we have $\min_{k \leq K} \left(S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k)) \right) = O(1/K)$.

Straightening Effect of Reflow (cont'd)

Proof.
We have

$$\mathbb{E}\left[\|X_1 - X_0\|^2\right] - \mathbb{E}\left[\|Z_1 - Z_0\|^2\right] = S(\boldsymbol{Z}) + V((X_0, X_1)). \tag{34}$$

Applying it to each rectification step yields

$$\mathbb{E}\left[\|Z_1^k - Z_0^k\|^2\right] - \mathbb{E}\left[\|Z_1^{k+1} - Z_0^{k+1}\|^2\right] = S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k)). \tag{35}$$

Take the sum over $k = 0, 1, \ldots, K$

$$\sum_{k=0}^{K}\left[\mathbb{E}\left[\|Z_1^k - Z_0^k\|^2\right] - \mathbb{E}\left[\|Z_1^{k+1} - Z_0^{k+1}\|^2\right]\right] = \sum_{k=0}^{K}\left[S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k))\right] \tag{36}$$

The left side is a telescoping sum, so only the first and last terms remain:

$$\mathbb{E}\left[\|Z_1^0 - Z_0^0\|^2\right] - \mathbb{E}\left[\|Z_1^{K+1} - Z_0^{K+1}\|^2\right] = \sum_{k=0}^{K}\left[S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k))\right] \tag{37}$$

Since $Z_0^0 = X_0$, $Z_1^0 = X_1$, we have

$$\mathbb{E}\left[\|X_1 - X_0\|^2\right] \geq \sum_{k=0}^{K}\left[S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k))\right] \tag{38}$$

which is exactly the conclusion stated in the theorem.

Straightening Effect of Reflow (cont'd)

Next, we show that $\min_{k \leq K} \left(S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k)) \right) = O(1/K)$.
From last page we have:

$$\sum_{k=0}^{K} \left[S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k)) \right] \leq \mathbb{E}\left[\|X_1 - X_0\|^2 \right] \tag{39}$$

Let $C := \mathbb{E}\left[\|X_1 - X_0\|^2 \right], a_k := S(\boldsymbol{Z}^{k+1}) + V((Z_0^k, Z_1^k))$, so

$$\sum_{k=0}^{K} a_k \leq C \tag{40}$$

Since minimum is less than the average,

$$\min_{0 \leq k \leq K} a_k \leq \frac{1}{K+1} \sum_{k=0}^{K} a_k \leq \frac{C}{K+1} = O(1/K) \tag{41}$$

Straightening Effect of Reflow (cont'd)

Finally, we will prove that (34) holds.

$$\mathbb{E}\left[\|X_1 - X_0\|^2\right] = \int_0^1 \mathbb{E}\left[\|X_1 - X_0\|^2\right] \mathrm{d}t \tag{42}$$

$$= \int_0^1 \mathbb{E}\left[\left\|(X_1 - X_0) - v^{\boldsymbol{X}}(X_t, t) + v^{\boldsymbol{X}}(X_t, t)\right\|^2\right] \mathrm{d}t \tag{43}$$

$$= \int_0^1 \mathbb{E}\left[\left\|(X_1 - X_0) - v^{\boldsymbol{X}}(X_t, t)\right\|^2\right] \mathrm{d}t \tag{44}$$

$$+ \int_0^1 \mathbb{E}\left[\|v^{\boldsymbol{X}}(X_t, t)\|^2\right] \mathrm{d}t$$

$$+ 2\int_0^1 \mathbb{E}\left[\left\langle (X_1 - X_0) - v^{\boldsymbol{X}}(X_t, t), v^{\boldsymbol{X}}(X_t, t)\right\rangle\right] \mathrm{d}t$$

where $v^{\boldsymbol{X}}(x, t) = \mathbb{E}[X_1 - X_0 \mid X_t = x]$.

Straightening Effect of Reflow (cont'd)

$$\mathbb{E}\left[\left\langle (X_1 - X_0) - v^{\boldsymbol{X}}(X_t, t), v^{\boldsymbol{X}}(X_t, t)\right\rangle\right] \tag{45}$$

$$= \mathbb{E}\left[\left\langle X_1 - X_0,\, v^{\boldsymbol{X}}(X_t, t)\right\rangle\right] - \mathbb{E}\left[\left\langle v^{\boldsymbol{X}}(X_t, t),\, v^{\boldsymbol{X}}(X_t, t)\right\rangle\right] \tag{46}$$

$$= \mathbb{E}\left[\left\langle X_1 - X_0,\, v^{\boldsymbol{X}}(X_t, t)\right\rangle\right] - \mathbb{E}\left[\|v^{\boldsymbol{X}}(X_t, t)\|^2\right] \tag{47}$$

$$\mathbb{E}\left[\left\langle X_1 - X_0,\, v^X(X_t, t)\right\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\left\langle X_1 - X_0,\, v^X(X_t, t)\right\rangle \mid X_t\right]\right] \tag{48}$$

$$= \mathbb{E}\left[\left\langle \mathbb{E}[X_1 - X_0 \mid X_t],\, v^X(X_t, t)\right\rangle\right] \tag{49}$$

$$= \mathbb{E}\left[\left\langle v^X(X_t, t),\, v^X(X_t, t)\right\rangle\right] \tag{50}$$

$$= \mathbb{E}\left[\|v^X(X_t, t)\|^2\right] \tag{51}$$

So we have

$$\mathbb{E}\left[\left\langle (X_1 - X_0) - v^{\boldsymbol{X}}(X_t, t), v^{\boldsymbol{X}}(X_t, t)\right\rangle\right] = 0 \tag{52}$$

Straightening Effect of Reflow (cont'd)

After simplification:

$$\mathbb{E}\left[\|X_1 - X_0\|^2\right] = \underbrace{\int_0^1 \mathbb{E}\left[\left\|(X_1 - X_0) - v^X(X_t, t)\right\|^2\right] \mathrm{d}t}_{V((X_0, X_1))} + \int_0^1 \mathbb{E}\left[\|v^X(X_t, t)\|^2\right] \mathrm{d}t \tag{53}$$

$$S(\boldsymbol{Z}) = \int_0^1 \mathbb{E}\left[\left\|(Z_1 - Z_0) - \dot{Z}_t\right\|^2\right] \mathrm{d}t \tag{54}$$

$$= \int_0^1 \mathbb{E}\left[\|Z_1 - Z_0\|^2 - 2\langle Z_1 - Z_0,\, \dot{Z}_t\rangle + \|\dot{Z}_t\|^2\right] \mathrm{d}t \tag{55}$$

$$= \mathbb{E}\left[\|Z_1 - Z_0\|^2\right] - 2\mathbb{E}\left[\left\langle Z_1 - Z_0,\, \int_0^1 \dot{Z}_t \mathrm{d}t\right\rangle\right] + \int_0^1 \mathbb{E}\left[\|\dot{Z}_t\|^2\right] \mathrm{d}t \tag{56}$$

$$= \mathbb{E}\left[\|Z_1 - Z_0\|^2\right] - 2\mathbb{E}\left[\|Z_1 - Z_0\|^2\right] + \int_0^1 \mathbb{E}\left[\|v^X(Z_t, t)\|^2\right] \mathrm{d}t \tag{57}$$

$$= \int_0^1 \mathbb{E}\left[\|v^X(X_t, t)\|^2\right] \mathrm{d}t - \mathbb{E}\left[\|Z_1 - Z_0\|^2\right] \tag{58}$$

So (34) is proven. □

Table of Contents

Reformulate the Transport Mapping Problem

- Let $\rho_0 :=$ prior density and $\rho_1 :=$ target density, both supported on $\Omega \subseteq \mathbb{R}^d$. The problem can be formulated as constructing a map $X_t : \Omega \to \Omega$ with $t \in [0, 1]$, such that

$$\text{if } x \sim \rho_0 \text{ then } X_t(x) \sim \rho_t \text{ with } \rho_{t=0} = \rho_0 \text{ and } \rho_{t=1} = \rho_1 \tag{59}$$

  where ρ_t is some density.

- Represent this map as the flow associated with the ODE

$$\dot{X}_t(x) = v_t(X_t(x)), \qquad X_{t=0}(x) = x. \tag{60}$$

  Here, v_t is the same as $v^{\boldsymbol{X}}$ in (10).

- This is equivalent to saying that $\rho_t(x)$ and $v_t(x)$ satisfies the continuity equation

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0 \text{ with } \rho_{t=0} = \rho_0 \qquad \text{and} \qquad \rho_{t=1} = \rho_1, \tag{61}$$

  and the problem becomes estimating v_t satisfying the equation.

Stochastic Interpolant

- Introduce a time-differentiable interpolant

$$I_t : \Omega \times \Omega \to \Omega \text{ with } I_{t=0}(x_0, x_1) = x_0 \text{ and } I_{t=1}(x_0, x_1) = x_1 \quad (62)$$

  E.g., linear interpolant $I_t(x_0, x_1) = tx_1 + (1-t)x_0$ in rectified flow.
- Given this interpolant, we then construct the stochastic process x_t by sampling independently x_0 from ρ_0 and x_1 from ρ_1, and passing them through I_t:

$$x_t = I_t(x_0, x_1), \qquad x_0 \sim \rho_0, \qquad x_1 \sim \rho_1 \qquad \text{independent.} \quad (63)$$

  We refer to the process x_t as a **stochastic interpolant**.

Table of Contents

Minimizer of Objective

Theorem 6

The stochastic interpolant x_t defined in (63) with $I_t(x_0, x_1)$ satisfying (62) has a probability density $\rho_t(x)$ that satisfies the continuity equation (61) with a velocity $v_t(x)$ which is the unique minimizer over $\hat{v}_t(x)$ of the objective

$$G(\hat{v}) = \mathbb{E}\left[|\hat{v}_t(I_t(x_0, x_1))|^2 - 2\partial_t I_t(x_0, x_1) \cdot \hat{v}_t(I_t(x_0, x_1))\right] \tag{64}$$

In addition, the minimum value of this objective is given by

$$G(v) = -\mathbb{E}\left[|v_t(I_t(x_0, x_1))|^2\right] = -\int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \rho_t(x) \mathrm{d}x \mathrm{d}t > -\infty \tag{65}$$

Equivalent Objective

- By definition of the stochastic interpolant x_t we can express its density $\rho_t(x)$ using the Dirac delta function as

$$\rho_t(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \delta\big(x - I_t(x_0, x_1)\big)\, \rho_0(x_0)\, \rho_1(x_1)\, \mathrm{d}x_0\, \mathrm{d}x_1. \tag{66}$$

- Take derivative w.r.t t using chain rule, we have:

$$\partial_t \rho_t(x) = -\int_{\mathbb{R}^d \times \mathbb{R}^d} \partial_t I_t(x_0, x_1) \cdot \nabla \delta\big(x - I_t(x_0, x_1)\big)\, \rho_0(x_0) \rho_1(x_1) \mathrm{d}x_0 \mathrm{d}x_1 \equiv -\nabla \cdot j_t(x) \tag{67}$$

  where

$$j_t(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \partial_t I_t(x_0, x_1)\, \delta\big(x - I_t(x_0, x_1)\big)\, \rho_0(x_0)\, \rho_1(x_1)\, \mathrm{d}x_0\, \mathrm{d}x_1. \tag{68}$$

- Introduce $v_t(x)$ via

$$v_t(x) = \begin{cases} j_t(x)/\rho_t(x) & \text{if } \rho_t(x) > 0, \\ 0 & \text{else} \end{cases} \tag{69}$$

  then we can write (67) as the continuity equation in (61).

Equivalent Objective (cont'd)

- Write (64) explicitly:

$$G(\hat{v}) = \int \mathrm{d}t \int \mathrm{d}x_0 \mathrm{d}x_1 \left(|\hat{v}_t(I_t(x_0, x_1))|^2 - 2\, \partial_t I_t(x_0, x_1) \cdot \hat{v}_t(I_t(x_0, x_1)) \right) \rho_0(x_0)\rho_1(x_1) \tag{70}$$

- Handle the two terms separately:

$$\int \mathrm{d}x_0 \, \mathrm{d}x_1 \; |\hat{v}_t(I_t(x_0, x_1))|^2 \; \rho_0(x_0)\rho_1(x_1) \tag{71}$$

$$= \int \mathrm{d}x_0 \, \mathrm{d}x_1 \int \mathrm{d}x \; |\hat{v}_t(I_t(x_0, x_1))|^2 \, \delta\big(x - I_t(x_0, x_1)\big)\rho_0(x_0)\rho_1(x_1) \tag{72}$$

$$= \int \mathrm{d}x \left|\hat{v}_t(x)\right|^2 \int \mathrm{d}x_0 \, \mathrm{d}x_1 \, \delta\big(x - I_t(x_0, x_1)\big)\rho_0(x_0)\rho_1(x_1) \tag{73}$$

$$= \int \mathrm{d}x \left|\hat{v}_t(x)\right|^2 \rho_t(x) \tag{74}$$

Similarly, we have:

$$\int \mathrm{d}x_0 \, \mathrm{d}x_1 \; (-2\, \partial_t I_t(x_0, x_1) \cdot \hat{v}_t(I_t(x_0, x_1))) \, \rho_0(x_0)\rho_1(x_1) = \int \mathrm{d}x \; (-2\hat{v}_t(x) \cdot j_t(x)) \tag{75}$$

Equivalent Objective (cont'd)

- Finally (64) becomes:

$$G(\hat{v}) = \int_0^1 \int_{\mathbb{R}^d} \left(|\hat{v}_t(x)|^2 \rho_t(x) - 2\hat{v}_t(x) \cdot j_t(x) \right) \mathrm{d}x \, \mathrm{d}t \tag{76}$$

- Consider the alternative objective which directly measures the distance between model and goal:

$$H(\hat{v}) = \int_0^1 \int_{\mathbb{R}^d} |\hat{v}_t(x) - v_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \tag{77}$$

$$= \int_0^1 \int_{\mathbb{R}^d} \left(|\hat{v}_t(x)|^2 \rho_t(x) - 2\hat{v}_t(x) \cdot j_t(x) + |v_t(x)|^2 \rho_t(x) \right) \mathrm{d}x \, \mathrm{d}t \tag{78}$$

- It follows that

$$G(\hat{v}) = H(\hat{v}) - \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t = H(\hat{v}) - \mathbb{E}\left[|v_t(I_t(x_0, x_1))|^2 \right] \tag{79}$$

- Clearly, (77) is equivalent to (9).

Table of Contents

Notations

Standard notation for function spaces:

- $C^1([0,1])$: The space of continuously differentiable functions from $[0,1]$ to $\mathbb{R}$.
- $(C^2(\mathbb{R}^d))^d$: The space of twice continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}^d$.
- $C_0^p(\mathbb{R}^d)$: The space of compactly supported functions from $\mathbb{R}^d$ to $\mathbb{R}$ that are continuously differentiable p times.
- Given a function $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ with value $b(t,x)$ at (t,x), $b \in C^1([0,1], (C^2(\mathbb{R}^d))^d)$ indicates that b is continuously differentiable in t for all $(t,x) \in [0,1] \times \mathbb{R}^d$ and that $b(t,\cdot)$ is an element of $(C^2(\mathbb{R}^d))^d$ for all $t \in [0,1]$.

(New) Stochastic Interpolant

- Given two probability density functions $\rho_0, \rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a stochastic interpolant between ρ_0 and ρ_1 is a stochastic process x_t defined as

$$x_t = I(t, x_0, x_1) + \gamma(t) z, \qquad t \in [0, 1], \tag{80}$$

where:

- $I \in C^2([0, 1]; (C^2(\mathbb{R}^d \times \mathbb{R}^d))^d)$ satisfies the boundary conditions $I(0, x_0, x_1) = x_0$ and $I(1, x_0, x_1) = x_1$, as well as

$$\exists C_1 < \infty : |\partial_t I(t, x_0, x_1)| \leq C_1 |x_0 - x_1| \qquad \forall (t, x_0, x_1) \in [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d. \tag{81}$$

- $\gamma : [0, 1] \to \mathbb{R}$ satisfies $\gamma(0) = \gamma(1) = 0$, $\gamma(t) > 0$ for all $t \in (0, 1)$, and $\gamma^2 \in C^2([0, 1])$.
- The pair (x_0, x_1) is drawn from a probability measure ν that marginalizes on ρ_0 and ρ_1, i.e.

$$\int \nu(x_0, x_1)\, \mathrm{d}x_1 = \rho_0(x_0), \int \nu(x_0, x_1)\, \mathrm{d}x_0 = \rho_1(x_1). \tag{82}$$

- z is a Gaussian random variable independent of (x_0, x_1), i.e. $z \sim \mathcal{N}(0, I_d)$ and $z \perp (x_0, x_1)$.

Details

- (81) restricts the speed of the interpolation trajectory, and prevents the trajectory from deviation.
- A simple choice of ν is the product measure $\nu(\mathrm{d}x_0, \mathrm{d}x_1) = \rho_0(x_0)\rho_1(x_1)\,\mathrm{d}x_0\mathrm{d}x_1$, where $x_0 \perp x_1$.
- We will see the advantage of the **additional term** $\gamma(t)z$, compared with (62).
- Another way to define the stochastic interpolant is via

$$x_t^{\mathrm{d}} = I(t, x_0, x_1) + N_t \tag{83}$$

where $N : [0, 1] \to \mathbb{R}^d$ is a zero-mean Gaussian stochastic process satisfying $N_{t=0} = N_{t=1} = 0$, so we only need to know the covariance matrix $\mathbb{E}[N_t N_t^\top]$ at each timestep.

Transport Equation

Theorem 7

The probability distribution of the stochastic interpolant x_t defined in (80) is absolutely continuous with respect to the Lebesgue measure at all times $t \in [0,1]$ and its time-dependent density $\rho(t)$ satisfies $\rho(0) = \rho_0$, $\rho(1) = \rho_1$, $\rho \in C^1([0,1]; C^p(\mathbb{R}^d))$ for any $p \in \mathbb{N}$, and $\rho(t,x) > 0$ for all $(t,x) \in [0,1] \times \mathbb{R}^d$. In addition, ρ solves the transport equation

$$\partial_t \rho + \nabla \cdot (b\rho) = 0, \tag{84}$$

where we defined the velocity

$$b(t,x) = \mathbb{E}[\dot{x}_t \mid x_t = x] = \mathbb{E}[\partial_t I(t, x_0, x_1) + \dot{\gamma}(t) z \mid x_t = x]. \tag{85}$$

This velocity is in $C^0([0,1]; (C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$, and such that

$$\forall t \in [0,1] : \int_{\mathbb{R}^d} |b(t,x)|^2 \rho(t,x) \mathrm{d}x < \infty. \tag{86}$$

Objective

Theorem 8

The velocity b defined in (85) is the unique minimizer in $C^0([0,1];(C^1(\mathbb{R}^d))^d)$ of the quadratic objective

$$\mathcal{L}_b[\hat{b}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{b}(t,x_t)|^2 - (\partial_t I(t,x_0,x_1) + \dot{\gamma}(t)z)\cdot\hat{b}(t,x_t)\right)\mathrm{d}t \tag{87}$$

where x_t is defined in (80) and the expectation is taken independently over $(x_0,x_1)\sim\nu$ and $z\sim\mathcal{N}(0,I_d)$.

This is a generalization of (64). An equivalent objective is
$\mathbb{E}\left(\frac{1}{2}|\hat{b}(t,x_t)|^2 - (\partial_t I(t,x_0,x_1) + \dot{\gamma}(t)z)\cdot\hat{b}(t,x_t)\right), \qquad t\in[0,1].$

Score

Theorem 9

The score of the probability density ρ specified in Theorem 7 is in $C^1([0,1];(C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$ and given by

$$s(t,x) = \nabla \log \rho(t,x) = -\gamma^{-1}(t)\mathbb{E}(z \mid x_t = x) \qquad \forall (t,x) \in (0,1) \times \mathbb{R}^d \tag{88}$$

In addition it satisfies

$$\forall t \in [0,1] : \int_{\mathbb{R}^d} |s(t,x)|^2 \rho(t,x) \mathrm{d}x < \infty, \tag{89}$$

and is the unique minimizer in $C^1([0,1];(C^1(\mathbb{R}^d))^d)$ of the quadratic objective

$$\mathcal{L}_s[\hat{s}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{s}(t,x_t)|^2 + \gamma^{-1}(t) z \cdot \hat{s}(t,x_t)\right) \mathrm{d}t \tag{90}$$

where x_t is defined in (80) and the expectation is taken independently over $(x_0, x_1) \sim \nu$ and $z \sim \mathcal{N}(0, I_d)$.

An equivalent objective is $\mathbb{E}\left(\frac{1}{2}|\hat{s}(t,x_t)|^2 + \gamma^{-1}(t) z \cdot \hat{s}(t,x_t)\right), \qquad t \in (0,1).$

Denoiser

- The quantity

$$\eta_z(t,x) = \mathbb{E}(z \mid x_t = x), \tag{91}$$

is defined as the denoiser.

- We can rewrite score on $t \in (0,1)$ (where $\gamma(t) > 0$) as:

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x). \tag{92}$$

- This denoiser is the minimizer of an equivalent expression to (90),

$$\mathcal{L}_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left(\frac{1}{2}\left|\hat{\eta}_z(t,x_t)\right|^2 - z \cdot \hat{\eta}_z(t,x_t)\right) \mathrm{d}t. \tag{93}$$

FP Equations

For any $\epsilon \in C^0([0,1])$ with $\epsilon(t) \geq 0$ for all $t \in [0,1]$, the probability density ρ specified in Theorem 7 satisfies:

- The forward Fokker-Planck equation

$$\partial_t \rho + \nabla \cdot (b_{\mathrm{F}} \rho) = \epsilon(t) \Delta \rho, \qquad \rho(0) = \rho_0, \tag{94}$$

  where we defined the forward drift

$$b_{\mathrm{F}}(t,x) = b(t,x) + \epsilon(t) s(t,x). \tag{95}$$

  (94) is solved **forward in time** from $t=0$ to $t=1$, and its solution for the initial condition $\rho(0) = \rho_0$ satisfies $\rho(1) = \rho_1$.

- The backward Fokker-Planck equation

$$\partial_t \rho + \nabla \cdot (b_{\mathrm{B}} \rho) = -\epsilon(t) \Delta \rho, \qquad \rho(1) = \rho_1, \tag{96}$$

  where we defined the backward drift

$$b_{\mathrm{B}}(t,x) = b(t,x) - \epsilon(t) s(t,x). \tag{97}$$

  (96) is solved **backward in time** from $t=1$ to $t=0$, and its solution for the final condition $\rho(1) = \rho_1$ satisfies $\rho(0) = \rho_0$.

To verify these two equations, just plug in definition of b_{F} or b_{B}, and note that $sp = (\nabla \log \rho)\rho = \nabla \rho$, so $\nabla \cdot (s\rho) = \nabla \cdot (\nabla \rho) = \Delta \rho$.

Velocity Field

- From (85) we can write

$$b(t,x) = v(t,x) - \dot{\gamma}(t)\gamma(t)s(t,x), \tag{98}$$

where s is the score given in (88) and we define the velocity field

$$v(t,x) = \mathbb{E}\big(\partial_t I(t,x_0,x_1) \mid x_t = x\big). \tag{99}$$

- The velocity field $v \in C^0([0,1];(C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$ and can be characterized as the unique minimizer of

$$\mathcal{L}_v[\hat{v}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{v}(t,x_t)|^2 - \partial_t I(t,x_0,x_1)\cdot\hat{v}(t,x_t)\right)\mathrm{d}t \tag{100}$$

Generative Models

At any time $t \in [0,1]$, the law of the stochastic interpolant x_t coincides with the law of the three processes X_t, X_t^{F}, and X_t^{B}, respectively defined as:

1. The solutions of the probability flow associated with the transport equation (84)

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = b(t, X_t), \tag{101}$$

   solved either forward in time from the initial data $X_{t=0} \sim \rho_0$ or backward in time from the final data $X_{t=1} = x_1 \sim \rho_1$.

2. The solutions of the forward SDE associated with the FPE (94)

$$\mathrm{d}X_t^{\mathrm{F}} = b_{\mathrm{F}}(t, X_t^{\mathrm{F}})\mathrm{d}t + \sqrt{2\epsilon(t)}\,\mathrm{d}W_t, \tag{102}$$

   solved forward in time from the initial data $X_{t=0}^{\mathrm{F}} \sim \rho_0$ independent of W.

3. The solutions of the backward SDE associated with the backward FPE (96)

$$\mathrm{d}X_t^{\mathrm{B}} = b_{\mathrm{B}}(t, X_t^{\mathrm{B}})\mathrm{d}t + \sqrt{2\epsilon(t)}\,\mathrm{d}W_t^{\mathrm{B}}, \qquad W_t^{\mathrm{B}} = -W_{1-t}, \tag{103}$$

   solved backward in time from the final data $X_{t=1}^{\mathrm{B}} \sim \rho_1$ independent of W^{B}. Alternatively, solution of (103) is given by $X_t^{\mathrm{B}} = Z_{1-t}^{\mathrm{F}}$ where Z_t^{F} satisfies

$$\mathrm{d}Z_t^{\mathrm{F}} = -b_{\mathrm{B}}(1-t, Z_t^{\mathrm{F}})\mathrm{d}t + \sqrt{2\epsilon(t)}\,\mathrm{d}W_t, \tag{104}$$

   solved forward in time from the initial data $Z_{t=0}^{\mathrm{F}} \sim \rho_1$ independent of W.

$x_t, X_t, X_t^{\mathrm{F}}$ and X_t^{B} are different stochastic processes, but their **laws all coincide with** $\rho(t)$ **at any time** $t \in [0,1]$.

Likelihood Control

Theorem 10

Let ρ denote the solution of the FP equation (94) with $\epsilon(t) = \epsilon > 0$. Given two velocity fields $\hat{b}, \hat{s} \in C^0([0,1]; (C^1(\mathbb{R}^d))^d)$, define

$$\hat{b}_{\mathrm{F}}(t,x) = \hat{b}(t,x) + \epsilon \hat{s}(t,x), \qquad \hat{v}(t,x) = \hat{b}(t,x) + \gamma(t)\dot{\gamma}(t)\hat{s}(t,x) \tag{105}$$

Let $\hat{\rho}$ denote the solution to the FP equation

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}_{\mathrm{F}} \hat{\rho}) = \epsilon \Delta \hat{\rho}, \qquad \hat{\rho}(0) = \rho_0. \tag{106}$$

Then,

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\epsilon} \left(\mathcal{L}_b[\hat{b}] - \min_{\hat{b}} \mathcal{L}_b[\hat{b}] \right) + \frac{\epsilon}{2} \left(\mathcal{L}_s[\hat{s}] - \min_{\hat{s}} \mathcal{L}_s[\hat{s}] \right), \tag{107}$$

and

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\epsilon} \left(\mathcal{L}_v[\hat{v}] - \min_{\hat{v}} \mathcal{L}_v[\hat{v}] \right) + \frac{\sup_{t \in [0,1]} (\gamma(t)\dot{\gamma}(t) - \epsilon)^2}{2\epsilon} \left(\mathcal{L}_s[\hat{s}] - \min_{\hat{s}} \mathcal{L}_s[\hat{s}] \right). \tag{108}$$

Diffusive Interpolants

Given two probability density functions $\rho_0, \rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a diffusive interpolant between ρ_0 and ρ_1 is a stochastic process x_t^{d} defined as

$$x_t^{\mathrm{d}} = I(t, x_0, x_1) + \sqrt{2a(t)} B_t, \qquad t \in [0, 1], \tag{109}$$

where:

- $I(t, x_0, x_1)$ is as in (80).
- $(x_0, x_1) \sim \nu$ with ν satisfying (82).
- $a(t) \in C^2([0, 1])$ with $a(0) > 0$ and $a(t) \geq 0$ for all $t \in [0, 1]$.
- B_t is a standard **Brownian bridge process**, independent of x_0 and x_1.

(109) has the same single-time statistics and time-dependent density $\rho(t, x)$ as the following stochastic interpolant:

$$x_t = I(t, x_0, x_1) + \sqrt{2a(t)t(1-t)}\, z \qquad \text{with} \qquad (x_0, x_1) \sim \nu,\ z \sim \mathcal{N}(0, I_d),\ (x_0, x_1) \perp z. \tag{110}$$

One-sided Interpolants for Gaussian ρ_0

Given a probability density function $\rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a one-sided stochastic interpolant **between** $\mathcal{N}(0, I_d)$ **and** ρ_1 is a stochastic process x_t^{os}

$$x_t^{\mathrm{os}} = \alpha(t)z + J(t, x_1), \qquad t \in [0, 1] \tag{111}$$

where:

- $J \in C^2([0,1]; C^2((\mathbb{R}^d)^d))$ satisfies the boundary conditions $J(0, x_1) = 0$ and $J(1, x_1) = x_1$.
- x_1 and z are independent random variables drawn from ρ_1 and $\mathcal{N}(0, I_d)$, respectively.
- $\alpha : [0,1] \to \mathbb{R}$ satisfies $\alpha(0) = 1$, $\alpha(1) = 0$, $\alpha(t) > 0$ for all $t \in [0, 1)$, and $\alpha^2 \in C^2([0,1])$.

By construction, $x_{t=0}^{\mathrm{os}} = z \sim \mathcal{N}(0, I_d)$ and $x_{t=1}^{\mathrm{os}} = x_1 \sim \rho_1$, so that the distribution of the stochastic process x_t^{os} **bridges** $\mathcal{N}(0, I_d)$ **and** ρ_1.

One-sided Interpolants for Gaussian ρ_0 (cont'd)

- (111) has the same density as the stochastic interpolant defined in (80) if we set $I(t, x_0, x_1) = J_t(x_1) + \delta(t)x_0$ and take $\delta^2(t) + \gamma^2(t) = \alpha^2(t)$, since $x_0 \sim \mathcal{N}(0, I_d)$.
- Velocity field b becomes

$$b(t, x) = \mathbb{E}(\dot{\alpha}(t)z + \partial_t J(t, x_1) \mid x_t^{\mathrm{os}} = x), \tag{112}$$

- Quadratic objective becomes

$$\mathcal{L}_b[\hat{b}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{b}(t, x_t^{\mathrm{os}})|^2 - (\dot{\alpha}(t)z + \partial_t J(t, x_1)) \cdot \hat{b}(t, x_t^{\mathrm{os}})\right) \mathrm{d}t. \tag{113}$$

  The expectation $\mathbb{E}$ is taken independently over $x_1 \sim \rho_1$ and $z \sim \mathcal{N}(0, I_d)$.
- The score is given by

$$s(t, x) = -\alpha^{-1}(t)\eta_z(t, x), \qquad \eta_z(t, x) = \mathbb{E}(z \mid x_t^{\mathrm{os}} = x), \tag{114}$$

- These functions are the unique minimizers of the objectives

$$\mathcal{L}_s[\hat{s}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{s}(t, x_t^{\mathrm{os}})|^2 + \gamma^{-1}(t)z \cdot \hat{s}(t, x_t^{\mathrm{os}})\right) \mathrm{d}t, \tag{115}$$

$$\mathcal{L}_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{\eta}_z(t, x_t^{\mathrm{os}})|^2 - z \cdot \hat{\eta}_z(t, x_t^{\mathrm{os}})\right) \mathrm{d}t. \tag{116}$$

Mirror Interpolants

Given a probability density function $\rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a mirror stochastic interpolant **between ρ_1 and itself** is a stochastic process x_t^{mir}

$$x_t^{\text{mir}} = K(t, x_1) + \gamma(t)z, \qquad t \in [0,1] \tag{117}$$

where:

- $K \in C^2([0,1]; C^2((\mathbb{R}^d)^d))$ satisfies the boundary conditions $K(0, x_1) = x_1$ and $K(1, x_1) = x_1$.
- x_1 and z are random variables drawn independently from ρ_1 and $\mathcal{N}(0, I_d)$, respectively.
- $\gamma : [0,1] \to \mathbb{R}$ satisfies $\gamma(0) = \gamma(1) = 0$, $\gamma(t) > 0$ for all $t \in (0,1)$, and $\gamma^2 \in C^1([0,1])$.

By construction, $x_{t=0}^{\text{mir}} = x_{t=1}^{\text{mir}} = x_1 \sim \rho_1$, so that the distribution of the stochastic process x_t^{mir} **bridges ρ_1 to itself**.

Mirror Interpolants (cont'd)

- Velocity field b becomes

$$b(t,x) = \mathbb{E}(\partial_t K(t,x_1) + \dot{\gamma}(t) z \mid x_t^{\mathrm{mir}} = x). \tag{118}$$

- Quadratic objective becomes

$$\mathcal{L}_b[\hat{b}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{b}(t,x_t^{\mathrm{mir}})|^2 - (\partial_t K(t,x_1) + \dot{\gamma}(t) z)\cdot \hat{b}(t,x_t^{\mathrm{mir}})\right) \mathrm{d}t. \tag{119}$$

The expectation $\mathbb{E}$ is taken independently over $x_1 \sim \rho_1$ and $z \sim \mathcal{N}(0, I_d)$.

- The score is given by

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x), \qquad \eta_z(t,x) = \mathbb{E}(z \mid x_t^{\mathrm{mir}} = x). \tag{120}$$

- These functions are the unique minimizers of the objectives

$$\mathcal{L}_s[\hat{s}] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{s}(t,x_t^{\mathrm{mir}})|^2 + \gamma^{-1}(t) z\cdot \hat{s}(t,x_t^{\mathrm{mir}})\right) \mathrm{d}t, \tag{121}$$

$$\mathcal{L}_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left(\frac{1}{2}|\hat{\eta}_z(t,x_t^{\mathrm{mir}})|^2 - z\cdot \hat{\eta}_z(t,x_t^{\mathrm{mir}})\right) \mathrm{d}t. \tag{122}$$

Spatially Linear Interpolants

Specialize the function I to be linear in both x_0 and x_1, i.e., we consider

$$x_t^{\text{lin}} = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z, \tag{123}$$

where:

- $(x_0, x_1) \sim \nu$.
- $z \sim \mathcal{N}(0, I_d)$ with $(x_0, x_1) \perp z$.
- $\alpha, \beta, \gamma^2 \in C^2([0,1])$ satisfy the conditions

$$\alpha(0) = \beta(1) = 1; \qquad \alpha(1) = \beta(0) = \gamma(0) = \gamma(1) = 0; \qquad \forall t \in (0,1): \ \gamma(t) > 0. \tag{124}$$

Spatially Linear Interpolants (cont'd)

- The velocity b and the score s can both be expressed in terms of the following three conditional expectations

$$\eta_0(t,x) = \mathbb{E}(x_0 \mid x_t^{\mathrm{lin}} = x), \qquad \eta_1(t,x) = \mathbb{E}(x_1 \mid x_t^{\mathrm{lin}} = x), \qquad \eta_z(t,x) = \mathbb{E}(z \mid x_t^{\mathrm{lin}} = x). \tag{125}$$

- Velocity field b becomes

$$b(t,x) = \dot{\alpha}(t)\eta_0(t,x) + \dot{\beta}(t)\eta_1(t,x) + \dot{\gamma}(t)\eta_z(t,x). \tag{126}$$

- The score is given by

$$s(t,x) = -\gamma^{-1}(t)\,\eta_z(t,x). \tag{127}$$

- η_0, η_1, η_z are the unique minimizers of the objectives

$$\mathcal{L}_{\eta_0}(\hat{\eta}_0) = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_0(t, x_t^{\mathrm{lin}})|^2 - x_0 \cdot \hat{\eta}_0(t, x_t^{\mathrm{lin}})\right] \mathrm{d}t, \tag{128}$$

$$\mathcal{L}_{\eta_1}(\hat{\eta}_1) = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_1(t, x_t^{\mathrm{lin}})|^2 - x_1 \cdot \hat{\eta}_1(t, x_t^{\mathrm{lin}})\right] \mathrm{d}t, \tag{129}$$

$$\mathcal{L}_{\eta_z}(\hat{\eta}_z) = \int_0^1 \mathbb{E}\left[\frac{1}{2}|\hat{\eta}_z(t, x_t^{\mathrm{lin}})|^2 - z \cdot \hat{\eta}_z(t, x_t^{\mathrm{lin}})\right] \mathrm{d}t. \tag{130}$$

The expectation is taken independently over $(x_0, x_1) \sim \nu$ and $z \sim \mathcal{N}(0, I_d)$.

Table of Contents

SMLD: Forward Process

- **SMLD** (Score Matching with Langevin Dynamics) uses a **variance exploding** SDE for the forward process:

$$dx_t = \sqrt{\frac{d[\sigma^2(t)]}{dt}}\, dw_t, \qquad x_0 \sim p_{\text{data}} \tag{131}$$

where $\sigma^2(t)$ is a non-decreasing noise schedule and $t \in [0, 1]$.

- This SDE has no drift term and can be solved as a variable-variance Brownian motion:

$$x_t = x_0 + \int_0^t \sqrt{\frac{d[\sigma^2(s)]}{ds}}\, dw_s \tag{132}$$

The increment term is a zero-mean Gaussian with covariance

$$\int_0^t \frac{d[\sigma^2(s)]}{ds} ds = \sigma^2(t) - \sigma^2(0) \tag{133}$$

Assuming $\sigma^2(0) = 0$, the variance is simply $\sigma^2(t)$.

- Therefore, the solution simplifies to

$$x_t = x_0 + \sigma(t)\, z \tag{134}$$

where $z \sim \mathcal{N}(0, I_d)$.

SMLD as Stochastic Interpolant

The SMLD forward process is a special case of the stochastic interpolant framework:

$$x_t = I(t, x_0, x_1) + \gamma(t)\, z \tag{135}$$

with

$$I(t, x_0, x_1) = x_0, \qquad \gamma(t) = \sigma(t) \tag{136}$$

- This is a **one-sided stochastic interpolant** from x_0 (data) to noise.
- At $t = 0$, $x_t = x_0$; as t increases, noise is gradually added.

Reverse Process in SMLD and Stochastic Interpolant

The forward process adds noise (no learning). The reverse process requires learning:

$$d\bar{x}_t = -\frac{d[\sigma^2(t)]}{dt} \nabla_x \log q_t(\bar{x}_t)\, dt + \sqrt{\frac{d[\sigma^2(t)]}{dt}}\, d\bar{w}_t \tag{137}$$

- The score function $\nabla_x \log q_t(x)$ is learned by a neural network $\mathbf{s}_\theta(x, t)$.
- Training pairs (x_0, x_t) are generated by the forward SDE.
- Loss: Denoising score matching

$$\min_\theta \mathbb{E}\left[\left\|\mathbf{s}_\theta(x_t, t) + \frac{x_t - x_0}{\sigma^2(t)}\right\|^2\right] \tag{138}$$

Connection to stochastic interpolant:

- Compare with the general framework: $I(t, x_0, x_1) = x_0$, $\gamma(t) = \sigma(t)$.
- The forward drift $b_F(t, x) = 0$, $\epsilon(t) = \frac{1}{2}\frac{d[\sigma^2(t)]}{dt}$, $b(t, x) = -\frac{1}{2}\frac{d[\sigma^2(t)]}{dt} s(t, x)$.
- The backward drift $b_B(t, x) = -\frac{d[\sigma^2(t)]}{dt} s(t, x)$, with $s(t, x) = \nabla_x \log q_t(\bar{x}_t)$.

DDPM: Forward Process

- **DDPM** (Denoising Diffusion Probabilistic Model) uses a **variance preserving** SDE for the forward process:

$$\mathrm{d}x_t = -\frac{1}{2}\beta(t)x_t\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}w_t, \tag{139}$$

  where $\beta(t)$ is the noise schedule, and $t \in [0, 1]$.

- The SDE above admits an analytical solution:

$$x_t = \alpha(t)x_0 + \sqrt{1 - \alpha(t)^2}\,z, \tag{140}$$

  where $z \sim \mathcal{N}(0, I_d)$ and

$$\alpha(t) = \exp\left(-\frac{1}{2}\int_0^t \beta(s)\mathrm{d}s\right). \tag{141}$$

DDPM as Stochastic Interpolant

The DDPM forward process is a special case of the stochastic interpolant framework:

$$x_t = I(t, x_0, x_1) + \gamma(t)\, z \tag{142}$$

with

$$I(t, x_0, x_1) = \alpha(t) x_0, \qquad \gamma(t) = \sqrt{1 - \alpha(t)^2} \tag{143}$$

- This is a **one-sided stochastic interpolant** from x_0 (data) to noise.
- At $t = 0$, $x_t = x_0$; as t increases, noise is gradually added.

Reverse Process in DDPM and Stochastic Interpolant

The forward process adds noise (no learning). The reverse process requires learning:

$$\mathrm{d}\bar{x}_t = \left(-\frac{1}{2}\beta(t)\,\bar{x}_t - \beta(t)\nabla_x \log q_t(\bar{x}_t)\right)\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\bar{w}_t \tag{144}$$

- The score function $\nabla_x \log q_t(x)$ is learned by a neural network $\mathbf{s}_\theta(x, t)$.
- Training pairs (x_0, x_t) are generated by the forward SDE.
- Loss: Denoising score matching

$$\min_\theta \mathbb{E}\left[\left\|\mathbf{s}_\theta(x_t, t) + \frac{1}{1-\alpha(t)^2}(x_t - \alpha(t)x_0)\right\|^2\right] \tag{145}$$

Connection to stochastic interpolant:

- Compare with the general framework: $I(t, x_0, x_1) = \alpha(t)x_0$, $\gamma(t) = \sqrt{1-\alpha(t)^2}$.
- The forward drift $b_{\mathrm{F}}(t, x) = -\frac{1}{2}\beta(t)x, \quad \epsilon(t) = \frac{1}{2}\beta(t)$.
- The backward drift $b_{\mathrm{B}}(t, x) = -\frac{1}{2}\beta(t)x - \beta(t)s(t, x)$, with $s(t, x) = \nabla_x \log q_t(x)$.

References I

Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. (2023).
Stochastic interpolants: A unifying framework for flows and diffusions.
arXiv preprint arXiv:2303.08797.

Albergo, M. S. and Vanden-Eijnden, E. (2022).
Building normalizing flows with stochastic interpolants.
arXiv preprint arXiv:2209.15571.

Baddar, M. (2024).
Rectified flows in a nutshell.
Medium blog post.
Published December 12, 2024.

References II

Hawley, S. H. (2024).
Flow with what you know.
Blog post.
Published November 13, 2024.

Liu, X., Gong, C., and Liu, Q. (2022).
Flow straight and fast: Learning to generate and transfer data with rectified flow.
arXiv preprint arXiv:2209.03003.

Monge, G. (1781).
Mémoire sur la théorie des déblais et des remblais.
Mem. Math. Phys. Acad. Royale Sci., pages 666–704.

Pishro-Nik, H. (2014).
Introduction to probability, statistics, and random processes.
Kappa Research LLC. [Online].

References III

Wikipedia contributors (2024).
Picard–Lindelöf theorem — Wikipedia, The Free Encyclopedia.
[Online].

Wikipedia contributors (2025).
Brownian bridge — Wikipedia, The Free Encyclopedia.
[Online].

Thank you!
Any questions?