Interpolation

$$\texttt{Deform}(\texttt{Rectify}(\{X_t\})) = \texttt{Rectify}(\texttt{Deform}(\{X_t\}))$$
$$\texttt{NaturalEulerRF}(\texttt{Transform}(\{X_t\})) = \texttt{Transform}(\texttt{NaturalEulerRF}(\{X_t\}))$$

Key Property: Equivariance to Deformation

- The interpolation X_t can be **any** smooth (deterministic or randomized) process connecting X_0 and X_1.
- Most methods are equivalent to using Affine Interpolations.

Affine Interpolations

$$X_t = \alpha_t X_1 + \beta_t X_0, \quad \text{with} \quad \alpha_0 = \beta_1 = 0, \alpha_1 = \beta_0 = 1.$$

- Straight trajectories only when $\alpha_t + \beta_t = 1$.
- Variance Preserving (VP) interpolation: $\alpha_t^2 + \beta_t^2 = 1$.
  - DDPM/DDIM [HJA20, SME20], VP-SDE/ODE [SSDK$^+$20]:

$$\alpha_t = \exp\left(-5(1-t)^2 - 0.05(1-t)\right), \quad \beta_t = \sqrt{1-\alpha_t^2},$$

  - Spherical linear interpolation (slerp) [ND21]:

$$\alpha_t = \sin(\frac{\pi}{2}t), \quad \beta_t = \cos(\frac{\pi}{2}t).$$

Straight Interpolation

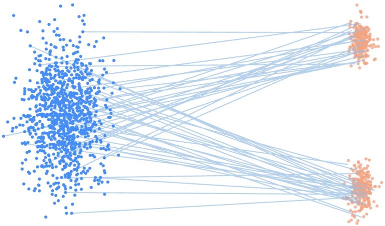
$$X_t = tX_1 + (1-t)X_0$$

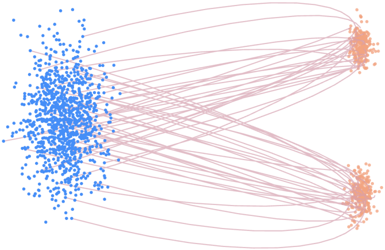
- Straight trajectories
- Not unit variance

Slerp Interpolation

$$X_t = \cos(\pi t/2)X_1 + \sin(\pi t/2)X_0$$

- Curved trajectories
- Unit variance





Equivalence of Deformed Interpolations

- Let $\{X_t\}$ and $\{X'_t\}$ be two interpolation processes from the same coupling, and $\{Z_t\}, \{Z'_t\}$ be their rectified flows:

$$\{Z_t\} = \texttt{Rectify}(\{X_t\}), \qquad \{Z'_t\} = \texttt{Rectify}(\{X'_t\}).$$

  Assume X_t and X'_t are related by an invertible pointwise transformation:

$$X'_t = \phi_t(X_{\tau_t}).$$

- Then the same transformation applies to their rectified flows:

$$Z'_t = \phi_t(Z_{\tau_t}).$$

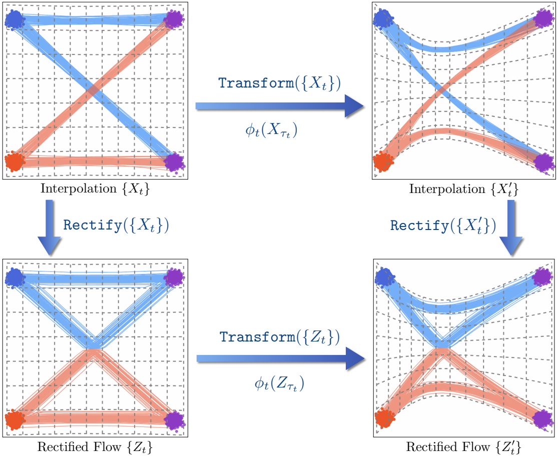
- Their velocities can be transformed by:

$$v'_t(x) = \partial_t \phi_t(\phi_t^{-1}(x)) + \left(\nabla \phi_t(\phi_t^{-1}(x))\right)^\top v_{\tau_t}(\phi_t^{-1}(x))\dot{\tau}_t.$$

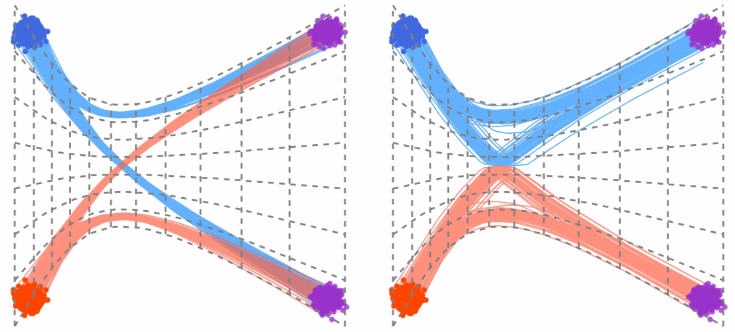
$$\texttt{Deform}(\texttt{Rectify}(\{X_t\})) = \texttt{Rectify}(\texttt{Deform}(\{X_t\}))$$

$$\texttt{Deform}(\texttt{Rectify}(\{X_t\})) = \texttt{Rectify}(\texttt{Deform}(\{X_t\}))$$

- Intuition: trajectory rewiring is equivariant under point-wise deformation.



Equivalence of Deformed Interpolations



Proof (assume $\phi_t(x_{\tau_t}) = \phi(x)$ for simplicity).

1) By definition, the velocity field v'_t of $\{Z'_t\} = \texttt{Rectify}(\{X'_t\})$ is

$$
\begin{aligned}
v'_t(x) &= \mathbb{E}\left[\dot{X}'_t \mid X'_t = x\right] \\
&= \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}\phi(X_t) \mid \phi(X_t) = x\right] \\
&= \mathbb{E}\left[\nabla\phi(X_t)^\top \dot{X}_t \mid X_t = \phi^{-1}(x)\right] \\
&= \nabla\phi(\phi^{-1}(x))^\top v_t(\phi^{-1}(x)). \qquad // v_t(x) = \mathbb{E}[\dot{X}_t | X_t = x]
\end{aligned}
$$

2) For $Z''_t = \phi(Z_t)$, we compute:

$$
\begin{aligned}
\dot{Z}''_t &= \frac{\mathrm{d}}{\mathrm{d}t}\phi(Z_t) = \nabla\phi(Z_t)^\top \dot{Z}_t = \nabla\phi(Z_t)^\top v_t(Z_t) \\
&= \nabla\phi(\phi^{-1}(Z''_t))^\top v_t(\phi^{-1}(Z''_t)) \\
&= v'_t(Z''_t),
\end{aligned}
$$

Hence, Z'_t and Z''_t coincides.

□

Affine Case: Time-Variable Scaling Transform

Pointwise Transform between Affine Interpolations

- Consider two affine interpolations from the same coupling (X_0, X_1):

$$X_t = \alpha_t X_1 + \beta_t X_0, \qquad X'_t = \alpha'_t X_1 + \beta'_t X_0.$$

- Then $\{X_t\}$ and $\{X'_t\}$ are time-wrapping and scaling of each other:

$$X'_t = \frac{1}{\gamma_t} X_{\tau_t}, \qquad \forall t \in [0, 1].$$

- The time-warp τ_t and rescaling factor γ_t are found by solving:

$$\frac{\alpha_{\tau_t}}{\beta_{\tau_t}} = \frac{\alpha'_t}{\beta'_t}, \qquad \gamma_t = \frac{\alpha_{\tau_t}}{\alpha'_t} = \frac{\beta_{\tau_t}}{\beta'_t}, \qquad \forall t \in (0, 1),$$

with $\tau_0 = 0, \quad \tau_1 = 1, \qquad \gamma_0 = \gamma_1 = 1.$

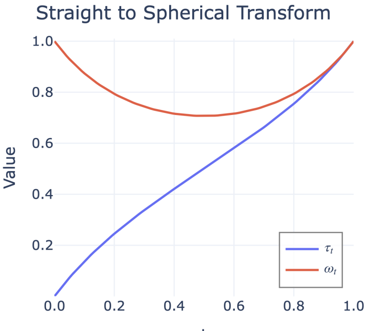
Example: Straight Flows to Affine Flows

Converting the straight interpolation $X_t = tX_1 + (1-t)X_0$ with $\alpha_t = t$ and $\beta_t = 1-t$ into another affine interpolation $X'_t = \alpha'_t X_1 + \beta'_t X_0$ gives:

$$\tau_t = \frac{\alpha'_t}{\alpha'_t + \beta'_t}, \quad \gamma_t = \frac{1}{\alpha'_t + \beta'_t}.$$

The velocity of affine flow:

$$v'_t(x) = \frac{\dot{\alpha}'_t \beta'_t - \alpha'_t \dot{\beta}'_t}{\alpha'_t + \beta'_t} v_{\tau_t}(\gamma_t x) \; + \; \frac{\dot{\alpha}'_t + \dot{\beta}'_t}{\alpha'_t + \beta'_t} x.$$



Key Property: Equivariance Under Numerical Methods

- Although different affine interpolations yield equivalent rectified flow ODEs, in practice we must solve them numerically.
- Euler method approximates trajectories with straight segments.
- **Curved segments** updates naturally arise in curved interpolations.

Natural Euler Method

Given an interpolation scheme $X_t = \mathtt{I}_t(X_0, X_1)$, define the update:

$$\hat{z}_{t_{i+1}} = \mathtt{I}_{t_{i+1}}\big(\hat{x}_{0|t_i}, \hat{x}_{1|t_i}\big),$$

where $\big(\hat{x}_{0|t_i}, \hat{x}_{1|t_i}\big)$ are expected noise and data at time t_i that match

$$\begin{cases} \hat{z}_{t_i} = \mathtt{I}_{t_i}\big(\hat{x}_{0|t_i}, \hat{x}_{1|t_i}\big), \\ v_{t_i}\big(\hat{z}_{t_i}\big) = \partial_t \mathtt{I}_t\big(\hat{x}_{0|t_i}, \hat{x}_{1|t_i}\big). \end{cases}$$

DDIM [SME20] as Natural Euler method

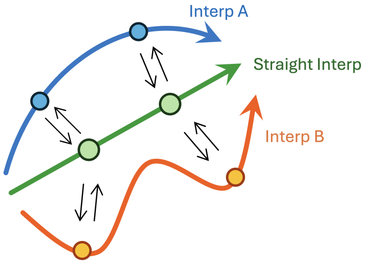
- DDIM employs a time-scaled spherical interpolation during training.
- DDIM inference scheme is an instance of the Natural Euler method.

$$\begin{aligned}\hat{Z}_{t+\epsilon} &= \alpha_{t+\epsilon} \cdot \hat{x}_{1|t}(\hat{Z}_t) + \beta_{t+\epsilon} \cdot \hat{x}_{0|t}(\hat{Z}_t) \\ &= \frac{\dot{\alpha}_t \beta_{t+\epsilon} - \alpha_{t+\epsilon} \dot{\beta}_t}{\dot{\alpha}_t \beta_t - \alpha_t \dot{\beta}_t} \hat{Z}_t + \frac{\alpha_{t+\epsilon} \beta_t - \alpha_t \beta_{t+\epsilon}}{\dot{\alpha}_t \beta_t - \alpha_t \dot{\beta}_t} v_t(\hat{Z}_t)\end{aligned}$$

Intuition: Natural Euler Method

- The Natural Euler update is equivalent to:

$$\texttt{Deform}^{-1}(\texttt{Euler}(\texttt{Deform}(Z)))$$



Equivalence of Natural Euler Trajectories

- Let $\{X_t\}$ and $\{X'_t\}$ be two interpolation processes from the same coupling, related by an invertible, time-dependent transform:

$$X'_t = \phi_t(X_{\tau_t}).$$

- Let $\{\hat{Z}_{t_i}\}$ and $\{\hat{Z}'_{t'_i}\}$ be their discrete natural Euler trajectories on grids $\{t_i\}$ and $\{t'_i\}$:

$$\hat{Z}_{t_i} = \texttt{NaturalEuler}(\{X_t\}, \{t_i\}), \quad \hat{Z}'_{t'_i} = \texttt{NaturalEuler}(\{X_t\}', \{t'_i\}).$$

- If the time grids align such that $\tau(t'_i) = t_i$, then the discrete trajectories match under the same transform:

$$\hat{Z}'_{t'_i} = \phi_{t'_i}(\hat{Z}_{t_i}), \quad \forall i.$$

- The final states always match, even if the paths differ:

$$\hat{Z}_1 = \hat{Z}'_1.$$

$$\texttt{NaturalEulerRF}(\texttt{Transform}(\{X_t\})) = \texttt{Transform}(\texttt{NaturalEulerRF}(\{X_t\}))$$

Example: Equivalence of Natural Euler method up to time rescaling

The following are equivalent:

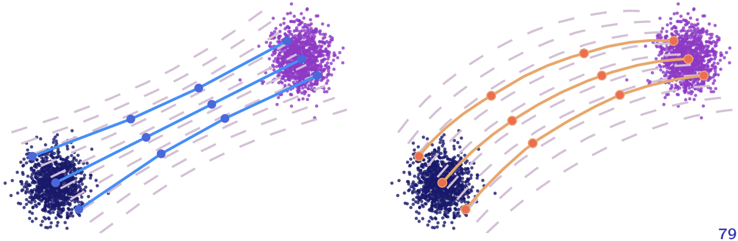
- `Natural Euler on` $X_t = \alpha_t X_1 + \beta_t X_0$, `on grid` $t_i = i/n$.
- `Vanilla Euler on` $X_t = tX_1 + (1-t)X_0$, `on grid` $t_i = \frac{\alpha_{i/n}}{\alpha_{i/n}+\beta_{i/n}}$.

Curved × Curved = Straight: DDIM is Straight RF

- DDIM: Natural Euler sampler under spherical interpolation.
- RF: (Natural) Euler sampler under straight interpolation.
- DDIM **"curved twice"**, but is equivalent to the standard Euler on straight RF with **rescaled time**.

$$\begin{aligned}\texttt{DDIM} &= \texttt{Deform} \circ \texttt{Euler} \circ \texttt{Deform}^{-1} \circ \texttt{Rectify} \circ \texttt{Deform}(\{X_t\}) \\ &= \texttt{Deform} \circ \texttt{Euler} \circ \texttt{Rectify}(\{X_t\}).\end{aligned}$$

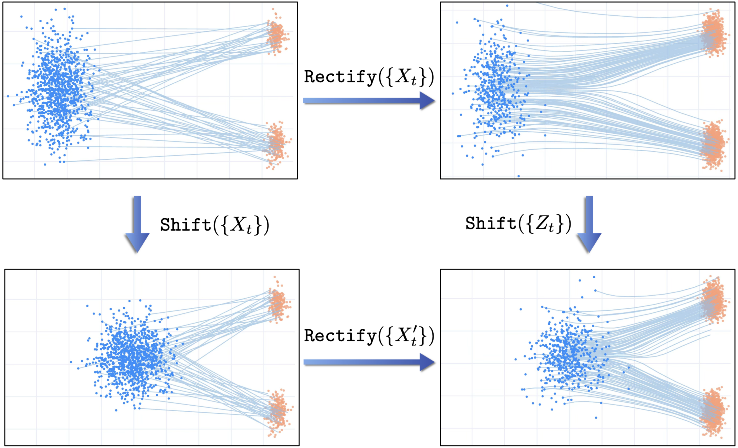
Different discretization time grids, same result.



Equivalent to Shifting and Scaling

- Applying shifting and scaling on the noise or data induces a transform on the RF trajectories.

$$\texttt{Shift}(\texttt{Rectify}(\{X_t\})) = \texttt{Rectify}(\texttt{Shift}(\{X_t\}))$$



Implication on Loss weights

- So far, we have assumed the model is perfectly trained.
- How does the choice of interpolation scheme affect training?

Loss weights during training

Learn a model $v_t(x;\theta)$ with an interpolation $X_t = \alpha_t X_1 + \beta_t X_0$, by

$$\mathcal{L}(\theta) = \int_0^1 \mathbb{E}\left[\omega_t \left\|\dot{X}_t - v_t(X_t;\theta)\right\|^2\right] dt,$$

It's equavalent to use another interpolation scheme $X'_t = \alpha'_t X_1 + \beta'_t X_0$ during training, but with the time-weighting and parameterization of:

$$\omega'_t = \frac{\gamma_t^2}{\dot{\tau}_t}\omega_{\tau_t}, \quad v'_t(x;\theta) = \frac{\dot{\tau}_t}{\gamma_t} v_{\tau_t}(\gamma_t x;\theta) - \frac{\dot{\gamma}_t}{\gamma_t} x$$

- Changing the interpolation scheme during training implicitly alters the loss weights and network parameterization.

Equivalence: Takeaways

Different methods lead to:

- different time-dependent loss weightings during training,
- different step size schedules during inference.

Open Questions

- How can we choose loss weights in a principled way?
- How should the inference scheme be determined during training?
- Related: Discussions on the equivalence between flow and diffusion [GHH+24, KAAL22, KG23, SPC+23], and on how to choose loss weights and time grids [Die24].