Introduction to Stochastic Interpolants

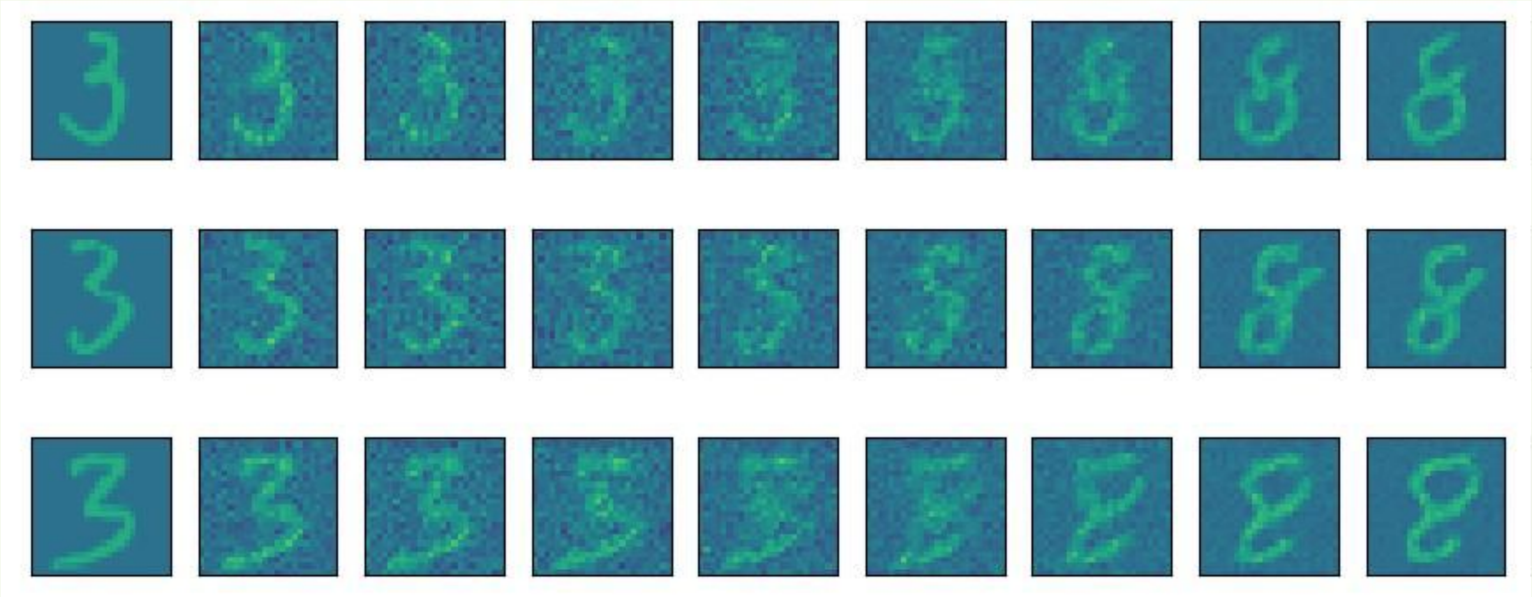
David Landry

Arches Weekly Meeting
2024-07-08

What is it

A generative model that uses Deep Neural Networks to learn the **flow** that goes from one distribution to another

This is attractive for **conditional generative problems**



Interpolating an MNIST 3 into and MNIST 8 using stochastic interpolants

Motivation

I study Stochastic Interpolants in the perspective of weather forecast postprocessing

Raw Ensemble Forecast -> Calibrated Weather Forecasts

Overview

- Build a basic SI
- Two sampling strategies
- Final notes

Let's build a Stochastic Interpolant

Build a stochastic interpolation model

Say we want to move samples from distribution ρ_0 to ρ_1

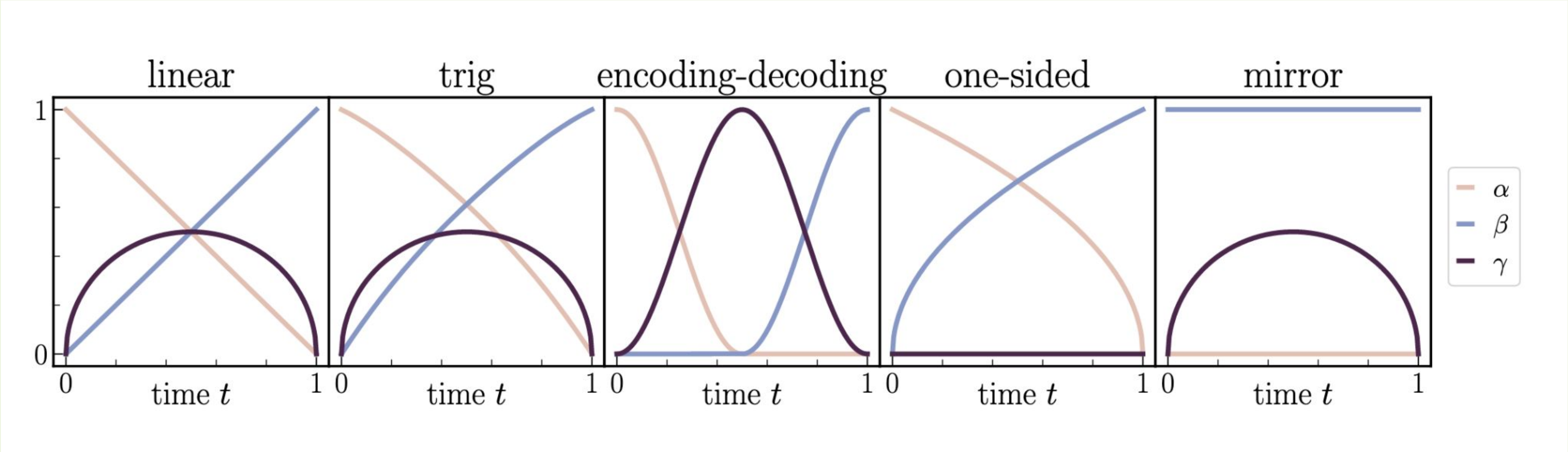
Interpolation function

If x_0 and x_1 are samples from ρ_0 and ρ_1, our interpolation function could be

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

where z is normal random noise.

α, β and γ are a trio that define our interpolation


linear
trig
encoding-decoding
one-sided
mirror
1
0
0
1
time t
α
β
γ

Approximate the probability flow

empirical loss

$$\frac{1}{n}\sum_{i=1}^{n} |b(t^i, x^i_{t^i})|^2 - 2b(t^i, x^i_{t^i}) \cdot (\partial_t I(t^i, x^i_0, x^i_1) + \dot{\gamma}(t^i) z^i)$$

→

sampler

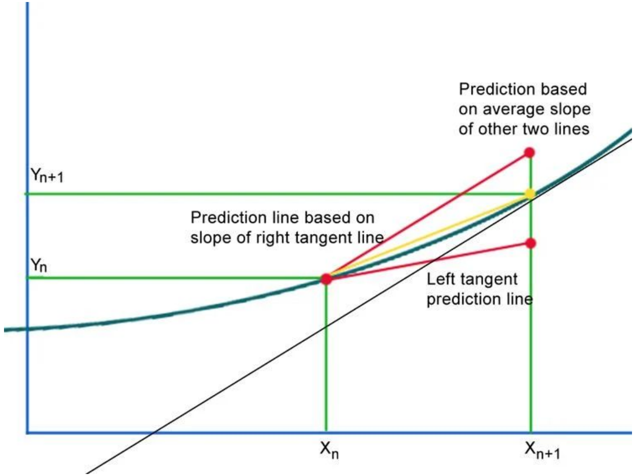
$$\frac{d}{dt} X_t = b(t, X_t)$$

A solver

Choose between a first-order and second order solver

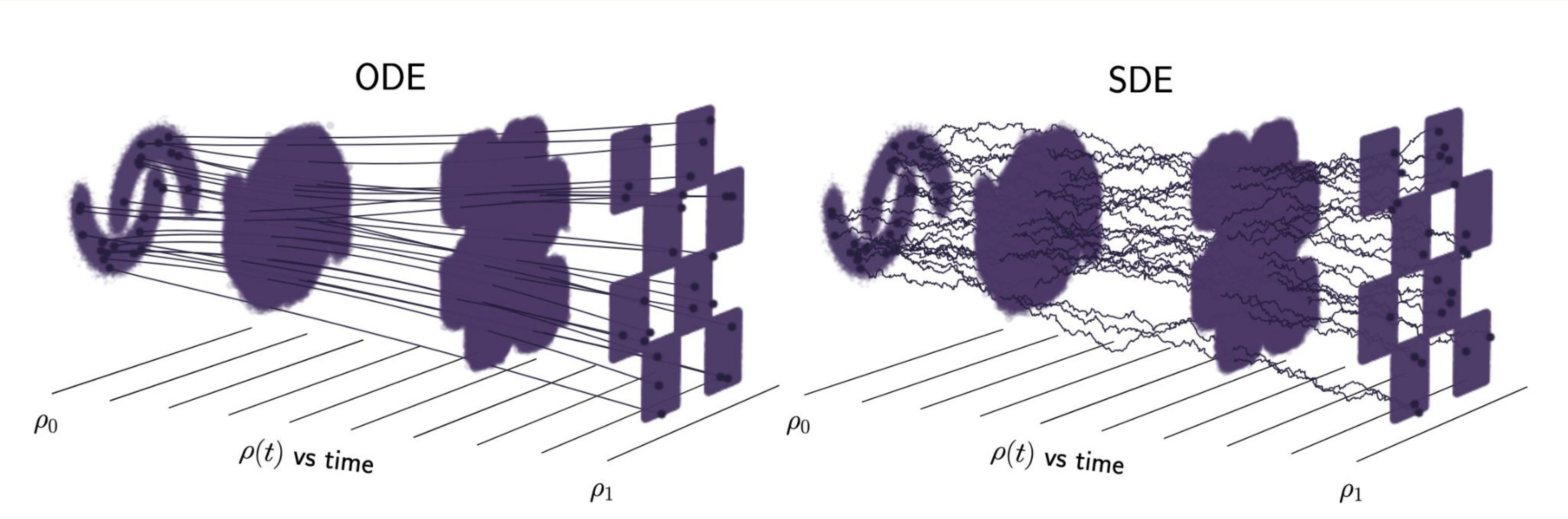
Denoising diffusion model community has custom solvers for some tasks

$$\text{Slope}_{\text{ideal}} = \frac{\Delta y}{h}$$
$$\Delta y = h(\text{Slope}_{\text{ideal}})$$
$$x_{i+1} = x_i + h, y_{i+1} = y_i + \Delta y$$
$$y_{i+1} = y_i + h\text{Slope}_{\text{ideal}}$$
$$y_{i+1} = y_i + \frac{1}{2}h(\text{Slope}_{\text{left}} + \text{Slope}_{\text{right}})$$
$$y_{i+1} = y_i + \frac{h}{2}(f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i)))$$



https://en.wikipedia.org/wiki/Heun's_method

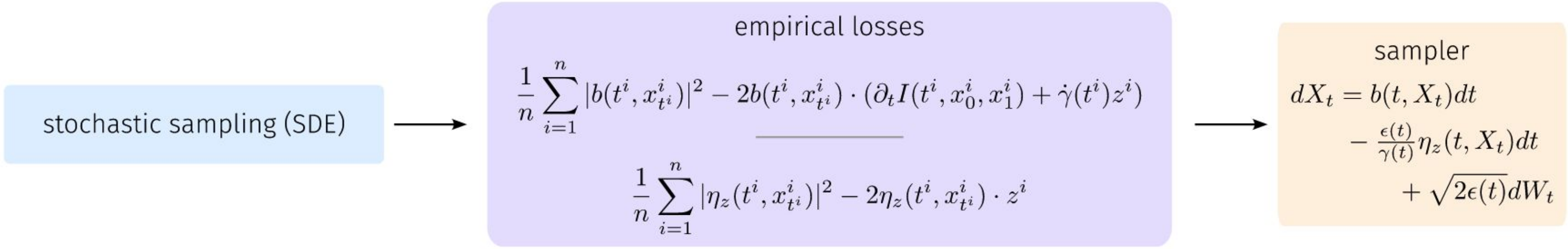
Two sampling strategies



Two sampling strategies: Ordinary Differential Equations vs Stochastic Differential Equations

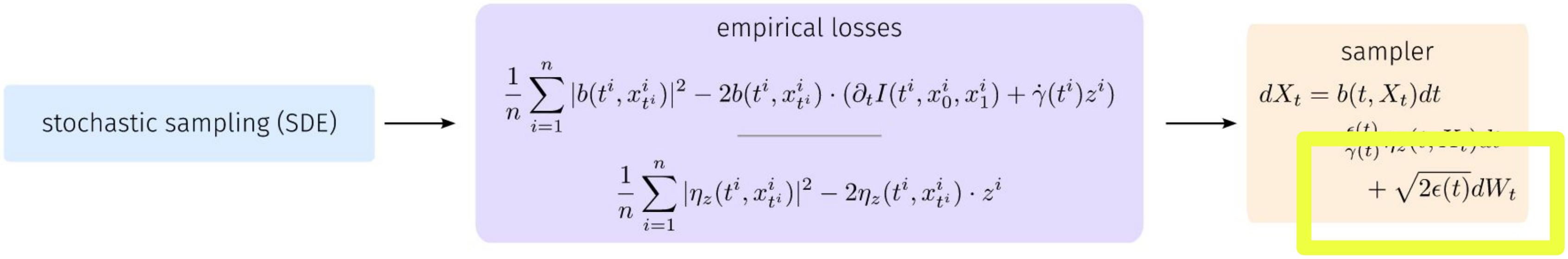
The SDE case

- We now need to learn two models fit on two different losses
- We also need a new function **ε(t)** which will determine how much noise is applied during sampling



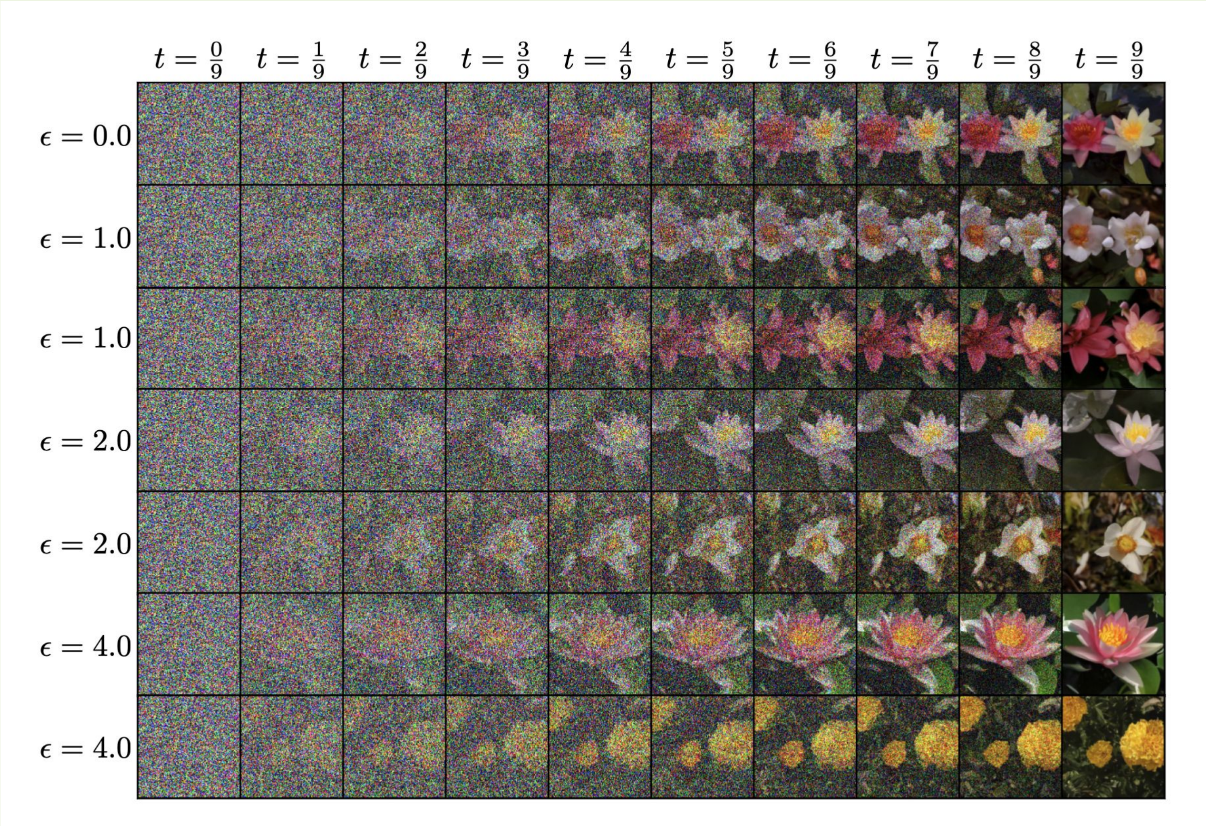
The SDE case

- We now need to learn two models fit on two different losses
- We also need a new function **ε(t)** which will determine how much noise is applied during sampling



Choice of ε

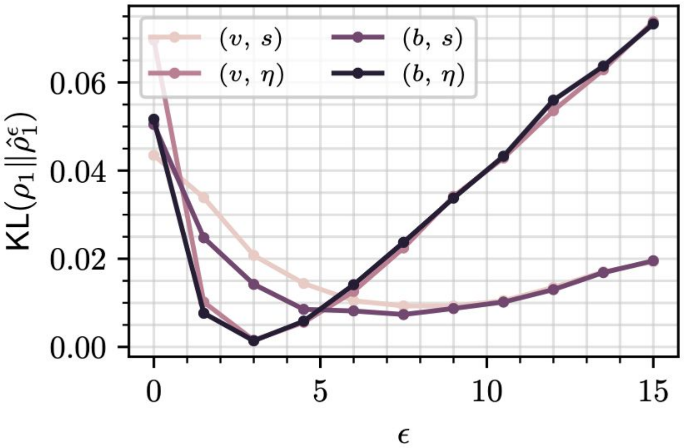
- Mathematically attractive to have it equal to $\gamma(t)$
  - There are some problems for sampling when γ is very close to zero.
- In practice the authors often set it to a constant
- If $\varepsilon(t)=0$, we recover the deterministic case
- When $\varepsilon(t)$ increases, we need more steps during sampling

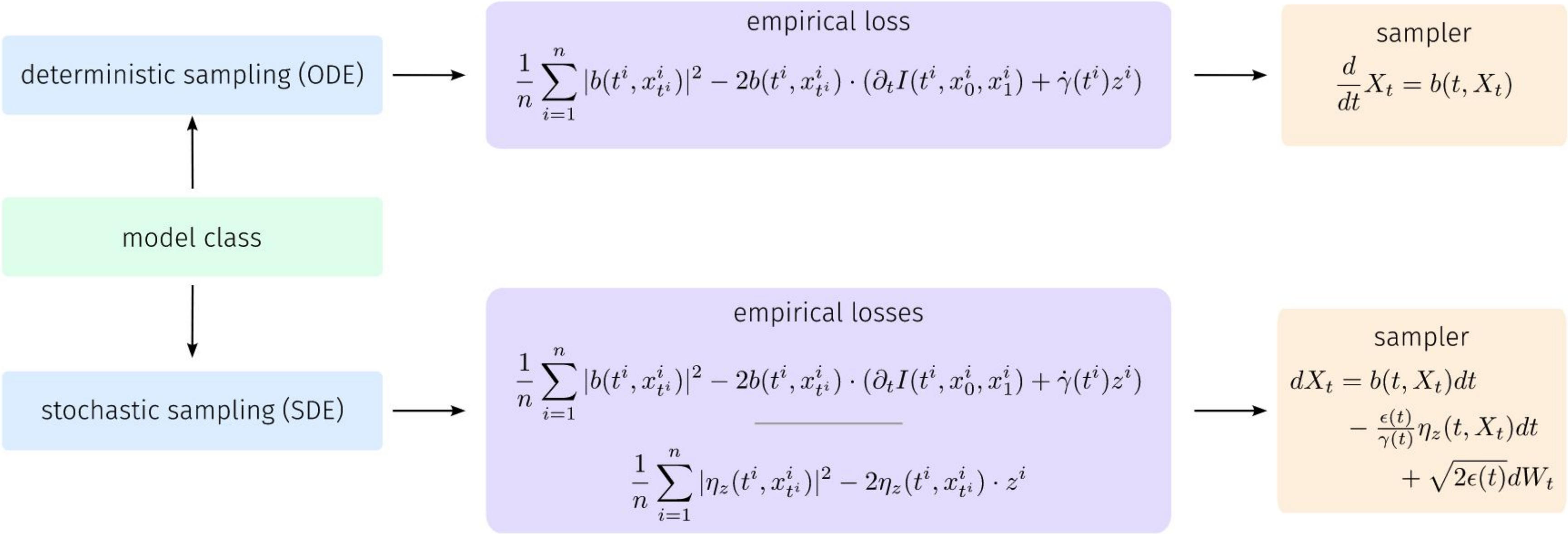


The choice of ε(t)

Should I use ODE or SDE?

- SDEs are a lot slower to sample (in the 1000s iterations) but seem to give better results

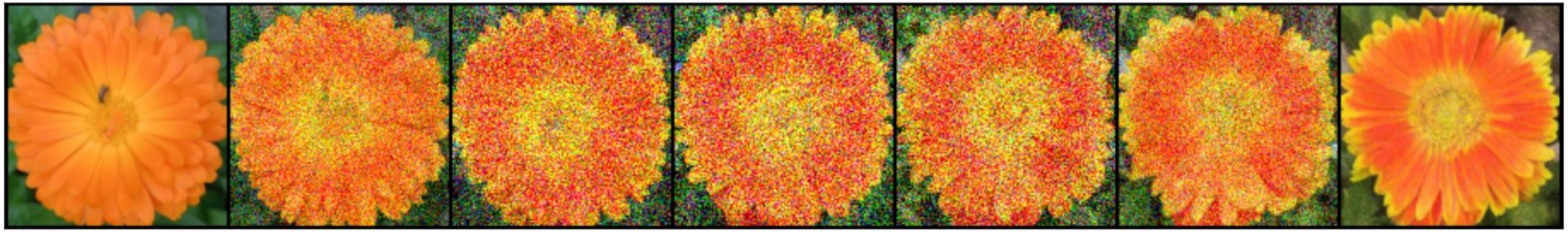


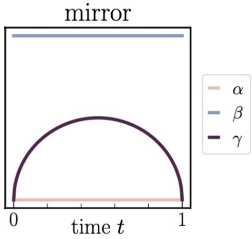

deterministic sampling (ODE)
model class
stochastic sampling (SDE)
empirical loss
$\frac{1}{n}\sum_{i=1}^{n} |b(t^i, x^i_{t^i})|^2 - 2b(t^i, x^i_{t^i}) \cdot (\partial_t I(t^i, x^i_0, x^i_1) + \dot{\gamma}(t^i) z^i)$
empirical losses
$\frac{1}{n}\sum_{i=1}^{n} |b(t^i, x^i_{t^i})|^2 - 2b(t^i, x^i_{t^i}) \cdot (\partial_t I(t^i, x^i_0, x^i_1) + \dot{\gamma}(t^i) z^i)$
$\frac{1}{n}\sum_{i=1}^{n} |\eta_z(t^i, x^i_{t^i})|^2 - 2\eta_z(t^i, x^i_{t^i}) \cdot z^i$
sampler
$\frac{d}{dt} X_t = b(t, X_t)$
sampler
$dX_t = b(t, X_t)dt - \frac{\epsilon(t)}{\gamma(t)} \eta_z(t, X_t)dt + \sqrt{2\epsilon(t)} dW_t$

Final notes

Mirror interpolants

What if ρ_0 and ρ_1 are the same distribution?

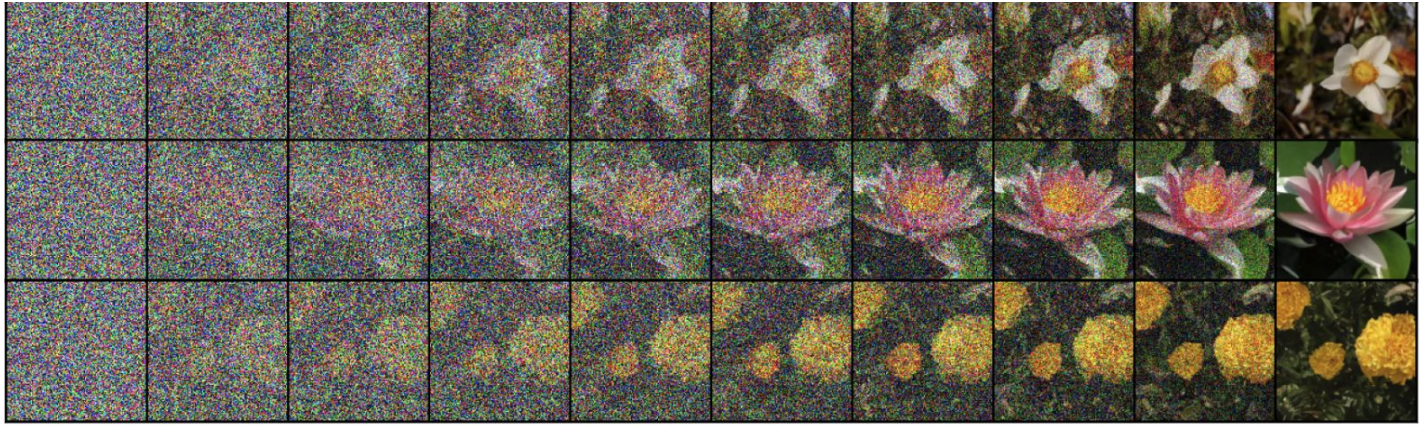


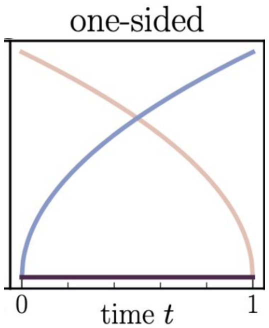


- Only need to learn η_z
- Motivation?

One-sided interpolant

- What if it is possible to sample from ρ_0?
- ρ_0 could be a normal distribution





Relationship with DDIMs and SBDMs

- Relationship to Denoising Diffusion Implicit Models (DDIMs) and Score-Based Diffusion models (SBDMs)
- We get something very close by using a one-sided interpolant with a normal ρ_0 distribution
- A difference is that the SI is a normal distribution at t=0, while diffusion models are normal at $t=\infty$
  - This poses mathematical difficulties which I will let smarter people argue about

References

1.

Albergo, M. S., Boffi, N. M. & Vanden-Eijnden, E. **Stochastic Interpolants: A Unifying Framework for Flows and Diffusions.** Preprint at https://doi.org/10.48550/arXiv.2303.08797 (2023).

Ma, N. et al. SiT: **Exploring Flow and Diffusion-based Generative Models with Scalable Interpolant Transformers.** (2024) doi:10.48550/ARXIV.2401.08740.