Stochastic Interpolants: A Unifying Framework for flows and diffusions

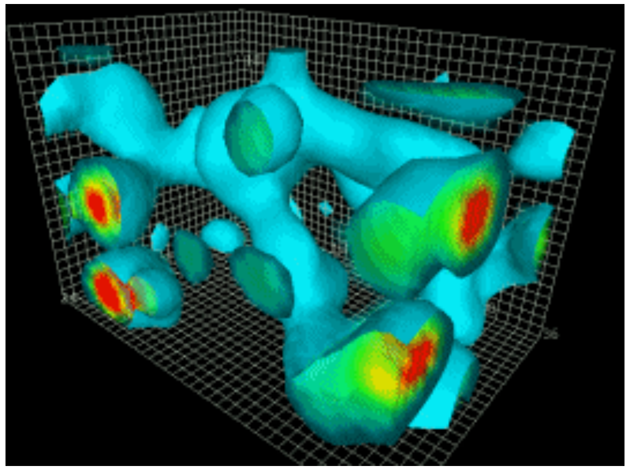
Michael Albergo
With Eric Vanden-Eijnden and Nicholas M. Boffi

Youth in High Dimensions, Trieste, Italy May 2023

Motivation + Research Directions

Research Interests

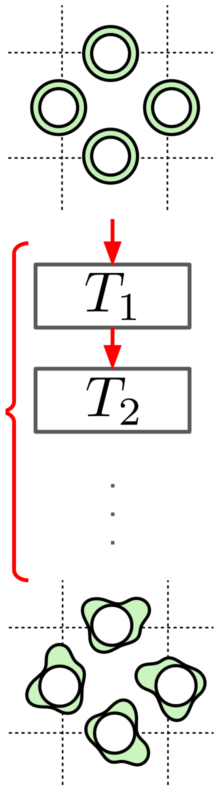
1 Computational approaches to field theory (high energy physics, condensed matter, …)

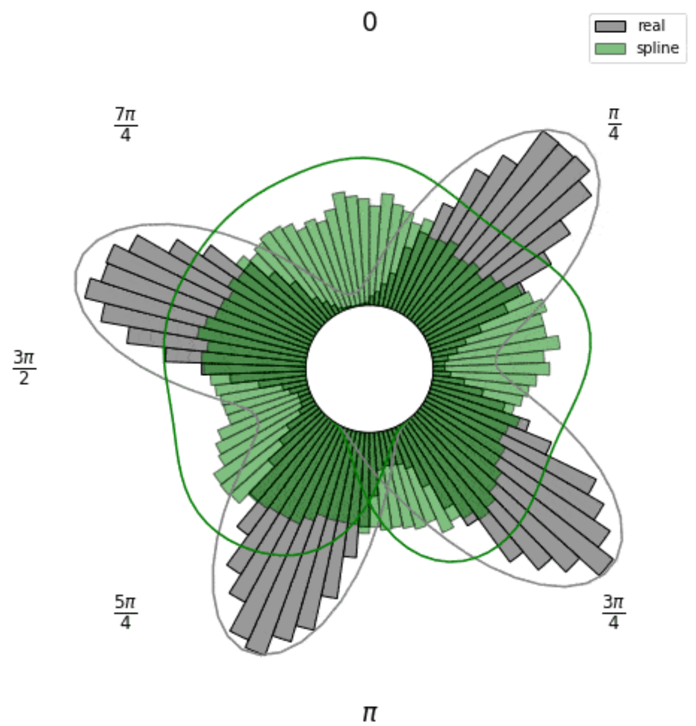


QCD vacuum energy density, aka QCD Lava Lamp

2 Machine learning techniques inspired by the underlying physics, as well as the demands of scientific computing

“Ab-initio AI”

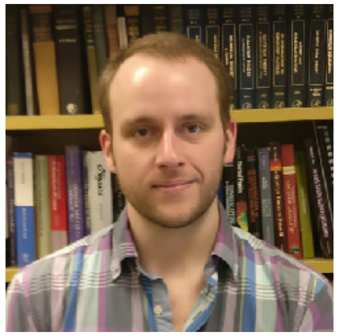




Thanks to collaborators and mentors

Massachusetts Institute of Technology


P. Shanahan


D. Hackett


F. Romero-


R. Abbott


D. Boyda

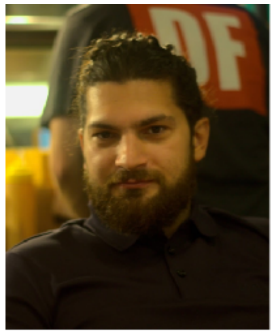

J. Urban

DeepMind


D. Rezende


S. Racanière


A. Razavi


A. Botev

NEW YORK UNIVERSITY


K. Cranmer


N. Boffi


R. Ranganath


E. Vanden-Eijnden

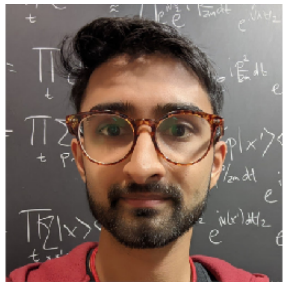

P. Lunts


M. Lindsey

FLATIRON INSTITUTE
Center for Computational Quantum Physics


A. Patel

UNIVERSITÄT BERN
AEC ALBERT EINSTEIN CENTER FOR FUNDAMENTAL PHYSICS


G. Kanwar

Agenda

Topic: density estimation and sampling with transport maps

↳ Motivation and background — the flow-based transport picture

Challenge: How to best learn expressive and scalable transport maps?

↳ Inspiration from score based diffusion

Stochastic Interpolants: Unifying flows and diffusions

↳ Unbiased generative modeling through either **deterministic** or **stochastic** dynamics

↳ ODE / SDE tradeoff, interpolant design

Problem Setup

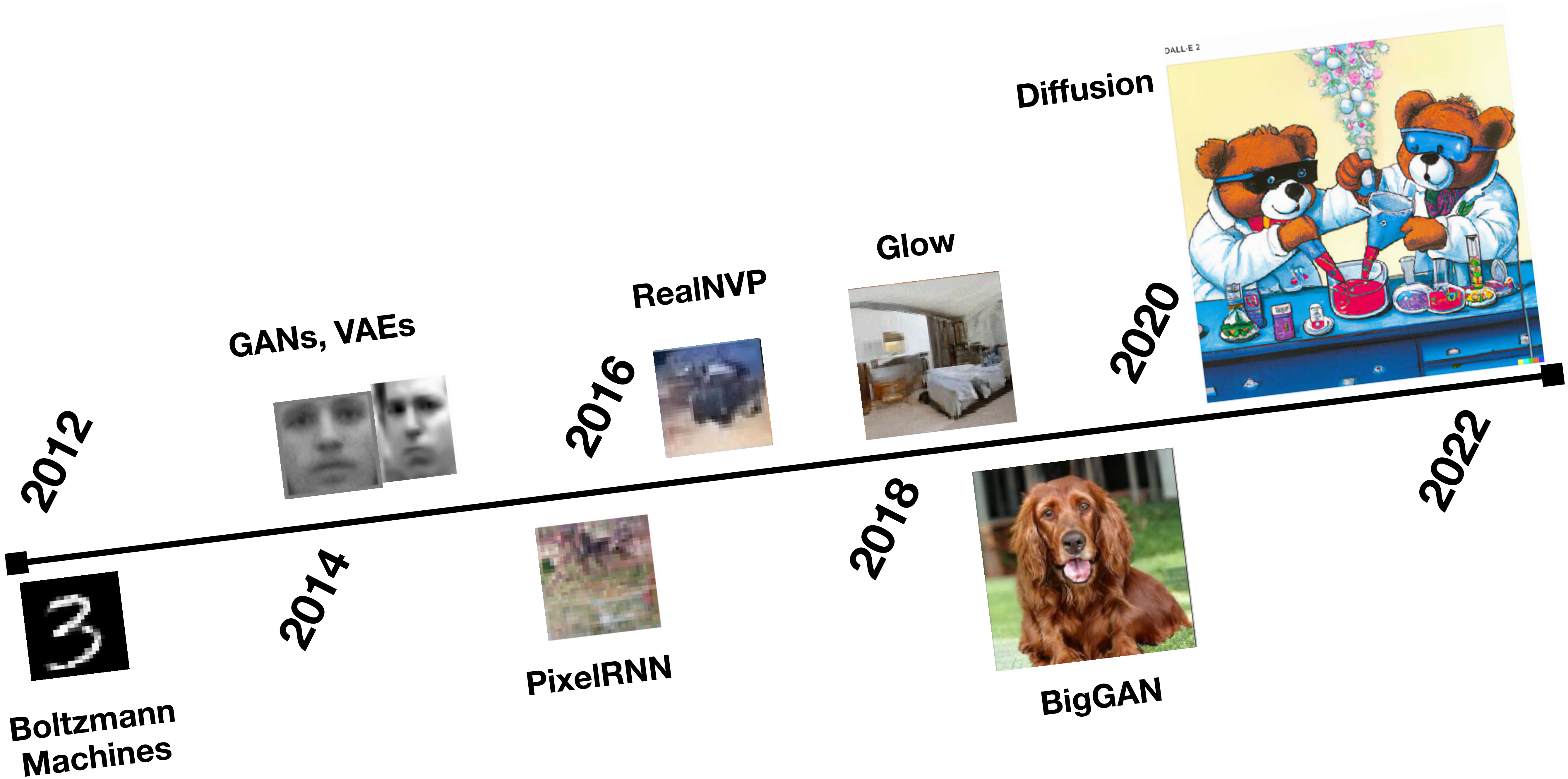
Goal: estimate the unknown *probability density function* $\rho_1 \in \mathscr{D}(\Omega)$ either through:

1. from sample data $\{x_i\}_{i=1}^n$
2. from query access to the unnormalized log likelihood

Problem Setup

Goal: estimate the unknown *probability density function* $\rho_1 \in \mathscr{D}(\Omega)$ either through:

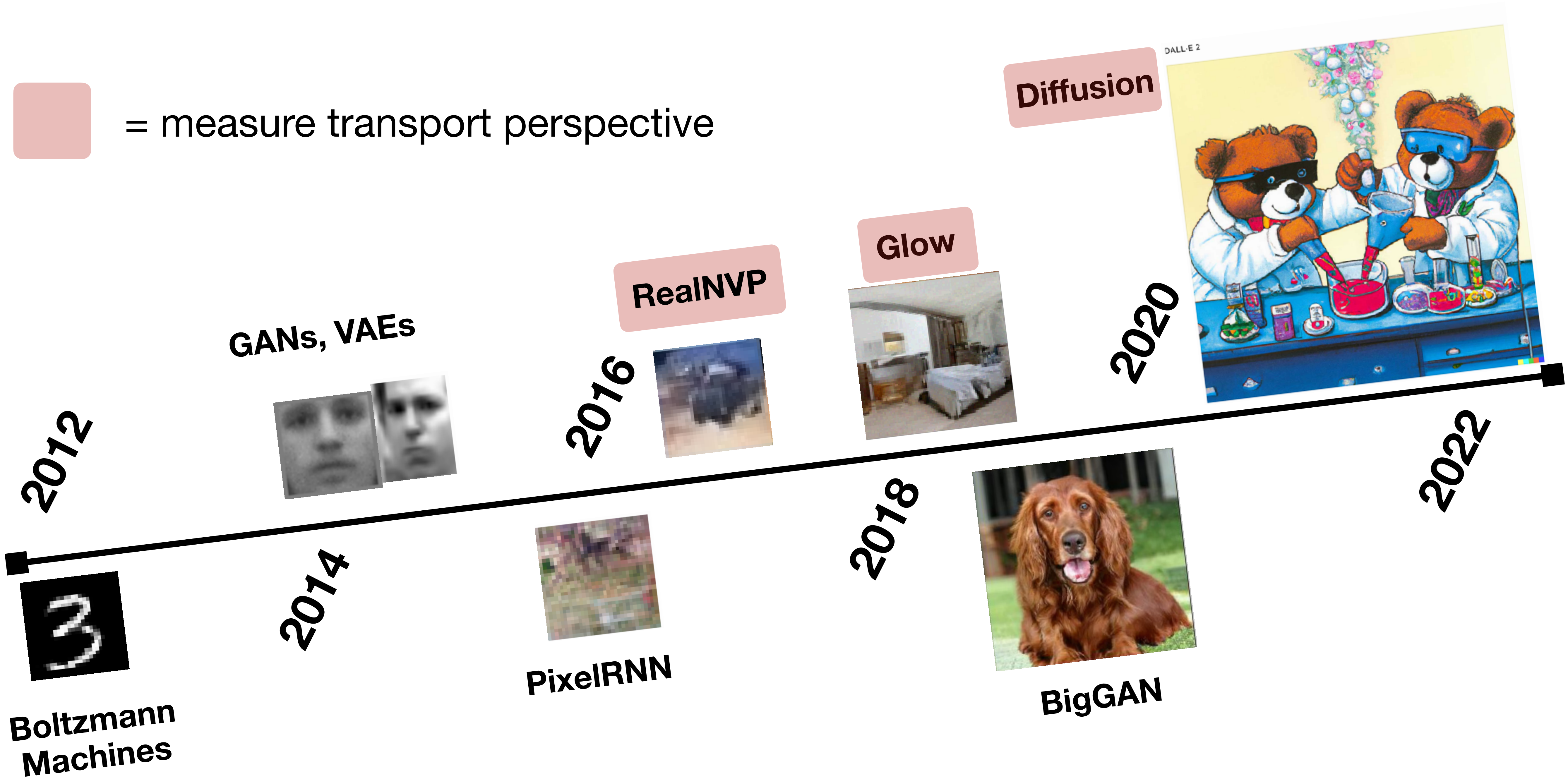
1. **from sample data $\{x_i\}_{i=1}^n$**
2. from query access to the unnormalized log likelihood



Problem Setup

Goal: estimate the unknown *probability density function* $\rho_1 \in \mathscr{D}(\Omega)$ either through:

1. **from sample data** $\{x_i\}_{i=1}^n$
2. from query access to the unnormalized log likelihood



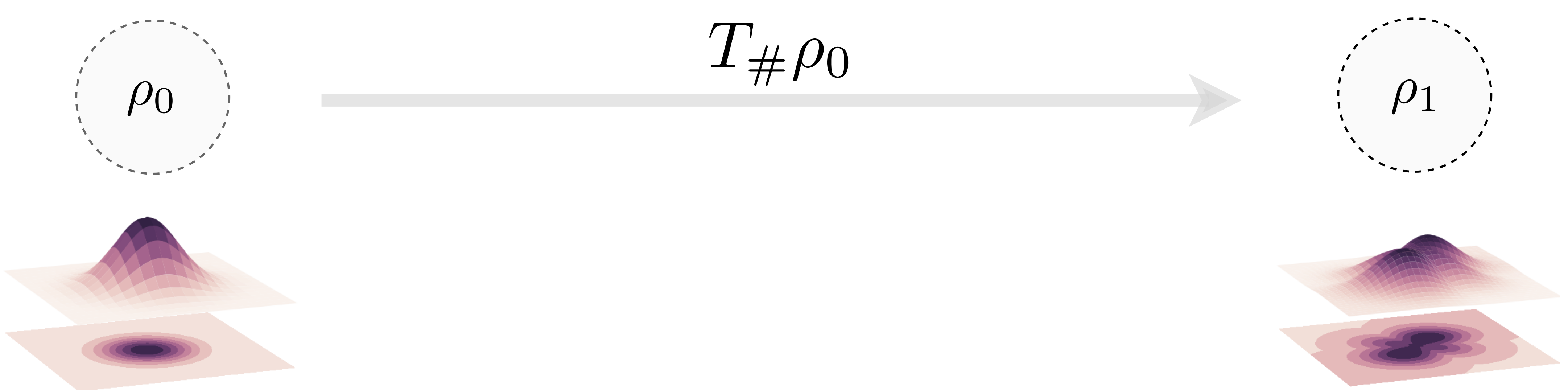
Problem Setup

Goal: estimate the unknown *probability density function* $\rho_1 \in \mathscr{D}(\Omega)$ either through:

1. **from sample data** $\{x_i\}_{i=1}^n$
2. from query access to the unnormalized log likelihood

The transport framework

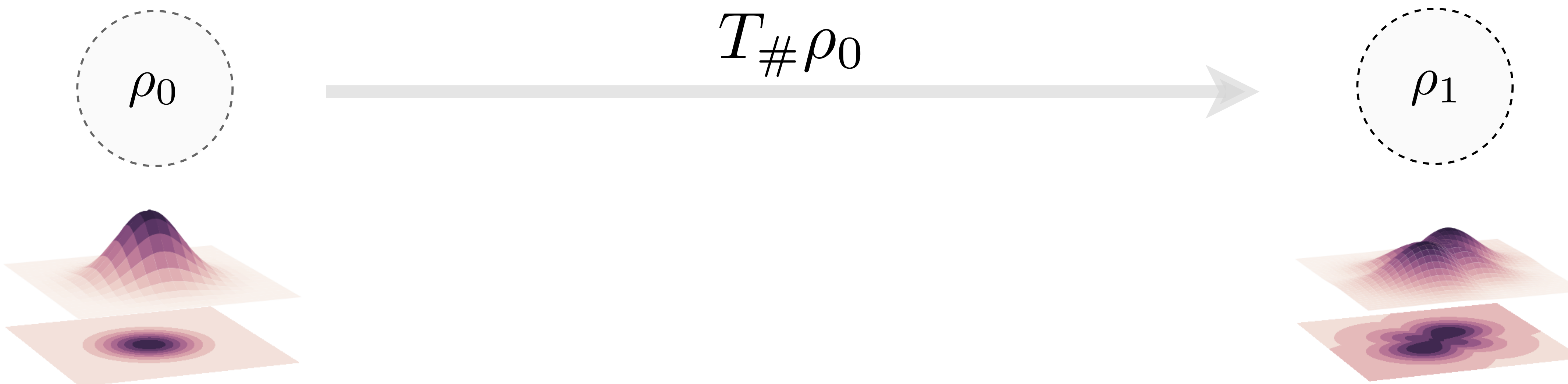
- Take a simple *base density* ρ_0 (e.g. Gaussian) and;
- Build a (reversible) map $T : \Omega \to \Omega$ such that the *pushforward of* ρ_0 *by* T *is* ρ_1*:* $T\sharp\rho_0 = \rho_1$



Likelihood under $\rho(1)$ given by: $\rho_1(x) = \rho_0(T^{-1}(x)) \det[\nabla T^{-1}(x)]$

The transport framework

- Build a (reversible) map $T : \Omega \to \Omega$ such that the *pushforward of* $\rho(0)$ *by* T *is* $\rho(1)$*:* $T\sharp\rho(0) = \rho(1)$



Likelihood: $\rho_1(x) = \rho_0(T^{-1}(x))\,\det[\nabla T^{-1}(x)]$

For parametric $\hat{T}(x)$ to be useful

- $\det[\nabla \hat{T}^{-1}(x)]$ to be **tractable**
- $\hat{T}(x)$ **maximally unconstrained**

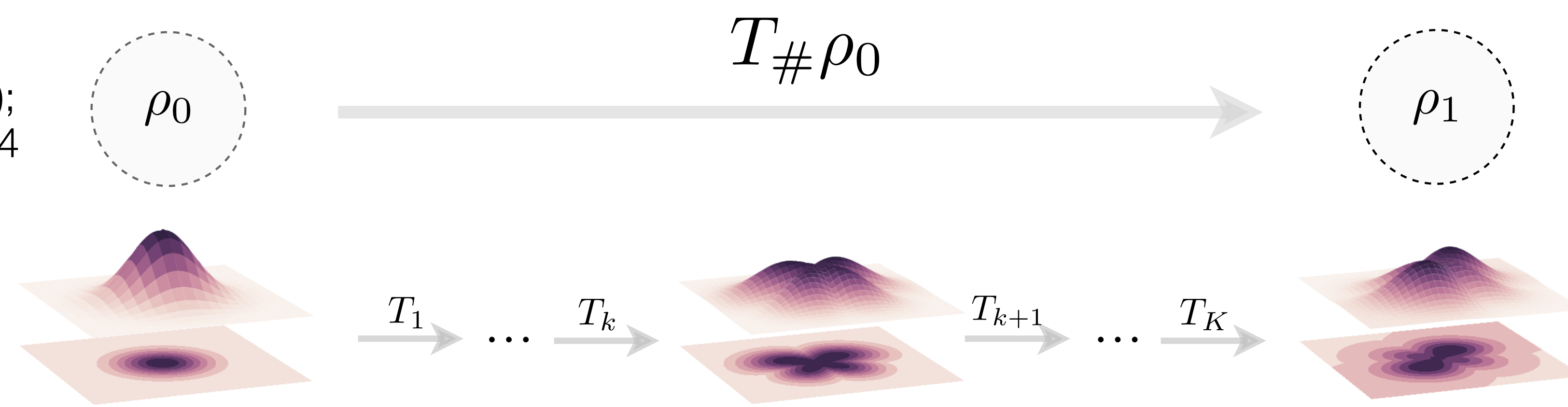
Tradeoff!

Series of discrete transforms

$\det[\nabla T^{-1}(x)]$ tractable, but too constrained?

T_k learned sequentially

Chen & Gopinath, NeurIPS 13 (2000);
Tabak & V.-E., Commun. Math. Sci. 8: 217-233 (2010);
Tabak & Turner, Comm. Pure App. Math LXVI, 145-164 (2013).



T_k structured invertible NNs

NICE: Dinh *et al.* arXiv:1410.8516 (2014);
Real NVP: Dinh *et al.* arXiv:1605.08803 (2016)
Rezende *et al.*, arXiv:1505.05770 (2015);
Papamakarios *et al.* arXiv:1912.02762 (2019); …

$k \to \infty$

T solution of ***continuous time flow***

FFJORD: Grathwohl *et al.* arXiv:1810.01367 (2018)

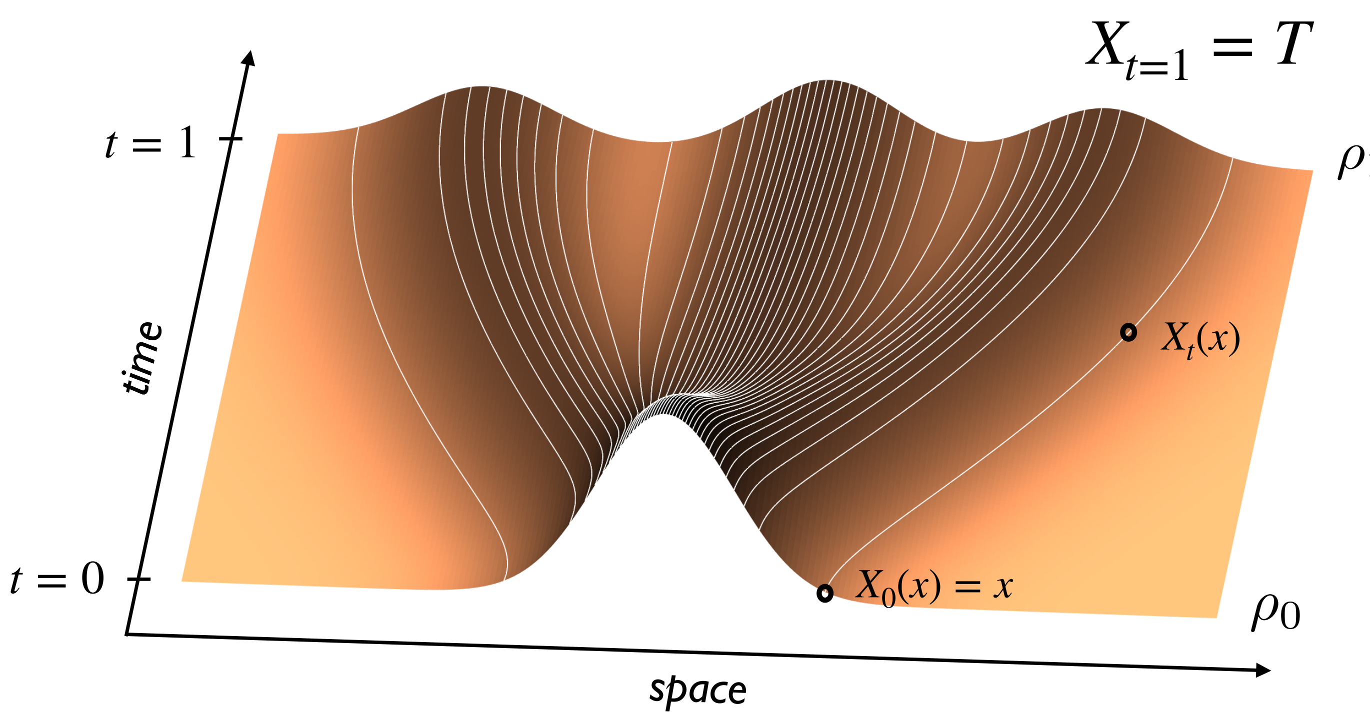
- $\det[\nabla T^{-1}(x)] \to \mathrm{Tr}[\frac{\partial b_t}{\partial x(t)}]$
- estimable via Skilling-Hutchinsion $O(D)$
- integrable with Neural ODEs

The continuous time picture

X_t flow map given by velocity field $b(t, x)$

$$X_{t=0}(x) = x \in \mathbb{R}^d$$

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

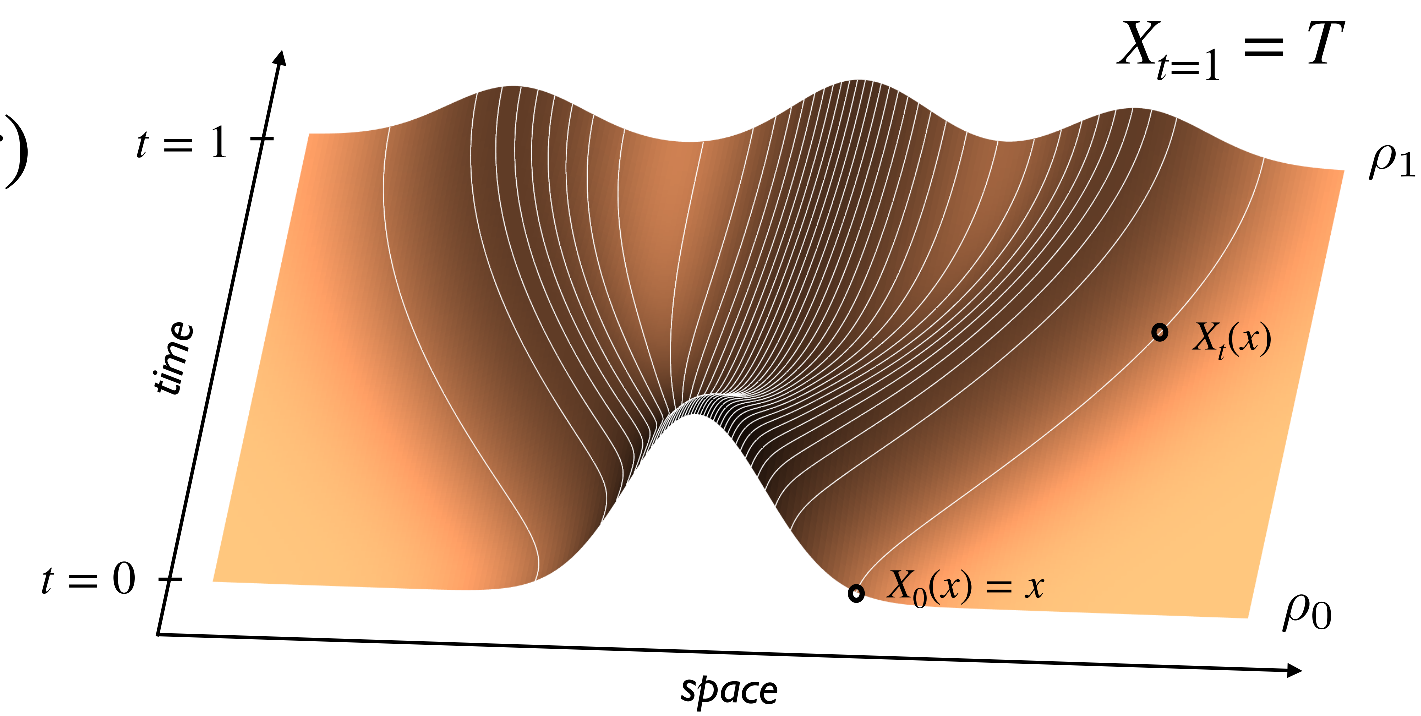


The continuous time picture

X_t flow map given by velocity field $b(t, x)$

$$X_{t=0}(x) = x \in \mathbb{R}^d$$

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



At the level of the of the distribution, how does $\rho(t, x)$ evolve?

Transport equation

$$\partial_t \rho(t, x) + \nabla \cdot \big(b(t, x)\rho(t, x)\big) = 0, \quad \rho(t = 0, \cdot) = \rho_0$$

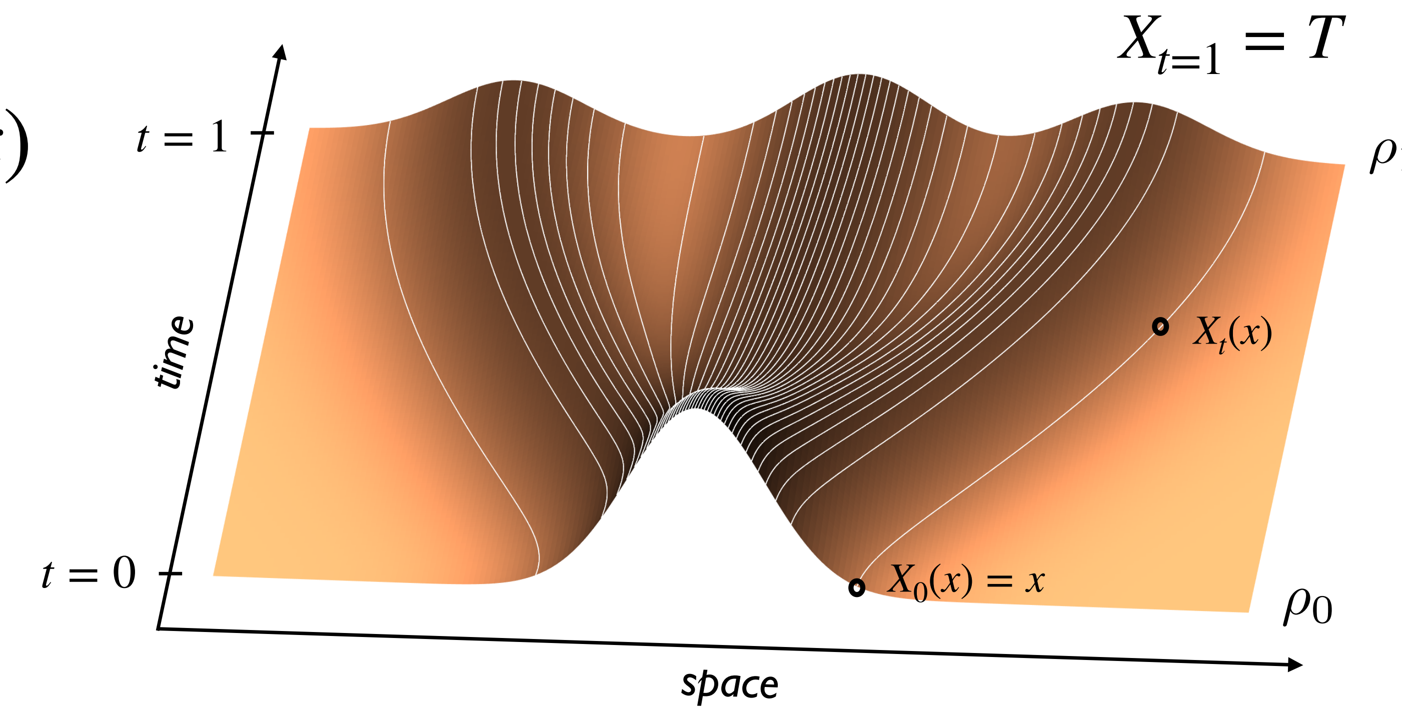
If $\rho(t)$ solves TE, **then** $\rho(t = 1, \cdot) = \rho_1$

The continuous time picture

X_t flow map given by velocity field $b(t, x)$

$$X_{t=0}(x) = x \in \mathbb{R}^d$$

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



At the level of the of the distribution, how does $\rho(t, x)$ evolve?

Transport equation

$$\partial_t \rho(t, x) + \nabla \cdot \big(b(t, x)\rho(t, x)\big) = 0, \quad \rho(t = 0, \cdot) = \rho_0$$

If $\rho(t)$ solves TE, **then** $\rho(t = 1, \cdot) = \rho_1$

Benamou-Brenier theory guarantees that $b(t, x)$ exists (assuming Lipschitz)

How to find a sufficient $b(t, x)$ to map ρ_0 to ρ_1?

Solving for $b(t, x)$ solves the transport

One approach: find $b(t, x)$ via maximum likelihood

FFJORD: Grathwohl *et al.* arXiv:1810.01367 (2018)

$$\rho(1, X_1(x)) = \rho_0(x) \exp\Big(- \int_0^1 \nabla \cdot b(t, X_t(x)) dt \Big)$$

$$\min_b KL(\rho_1 \,||\, \rho(1)) = \min \mathbb{E}_{\rho_1}\Big[\log \frac{\rho_1(x)}{\rho(1, x)}\Big]$$

$$= \min - \mathbb{E}_{\rho_1}\Big[\log \rho(1, x)\Big] + C$$

- $b(t, x)$ parametrized as neural network
- adjoint method (Neural ODE) allows for gradient wrt parameters of b

Loss involves integrating the ODE

Many paths from ρ_0 to ρ_1

Is there a simpler paradigm for learning $b(t, x)$?

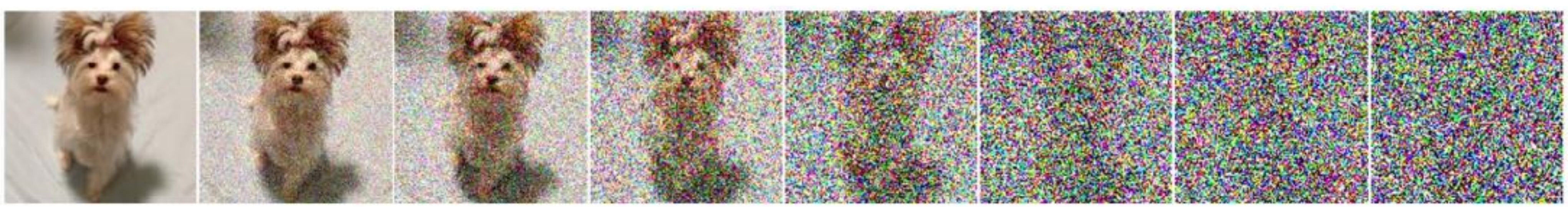
Inspiration: Score-based diffusion

Song et al. arXiv:2011.13456 (2021);
Sohl-Dickstein et al arXiv:1503.03585 (2021);
Hyvärinen JMLR **6** (2005);
Vincent, Neural Comp. **23**, 1661 (2011)

"A brain riding a rocket ship headed toward the moon." Imagen, Saharia et al 2205.11487

Map $x_1 \sim \rho_1$ to Gaussian ρ_0 via Ornstein-Uhlenbeck (OU) process

$$dX_t = -X\,dt + \sqrt{2}\,dW_t, \quad X_0 = x_1$$



SDE $$dX_t^B = -X_t\,dt + \nabla \log \rho(t, X_t)\,dt + \sqrt{2}\,dW_t, \quad X_0 = x_0$$

ODE $$b(t,x) = x - \nabla \log \rho(t,x)$$

Access to the score $s(t,x) = \nabla \log \rho(t,x)$ allows one to simulate the reverse process as a generative model

Inspiration: Score-based diffusion

Why does it work so well?

$$dX_t^B = -X_t\,dt + \nabla \log \rho(t, X_t)\,dt + \sqrt{2}\,dW_t$$

- Data available from $\rho(t, x)$ **for any** t: $X_t = xe^{-t} + \sqrt{2}\int_0^t e^{-t+s} dW_s$
- By choosing a path in the space of measures, turns generative modeling into a **regression problem**

score matching
Hyvarinen 2005

$$s(t,x) = \underset{\hat{s}(t,x)}{\text{argmin}} \int |\hat{s}(t,x) - \nabla \log \rho(t,x)|^2 \rho(t,x) dx$$

$$= \underset{\hat{s}(t,x)}{\text{argmin}} \int \left(|\hat{s}(t,x)|^2 + 2\nabla \cdot \hat{s}(t,x) \right) \rho(t,x) dx$$

Limitations?

- Requires one of the endpoints of the transport to be Gaussian
- Requires $t \to \infty$ in noising interval. Truncation over $t \in [0,T] \quad T \gg 1$ introduces **bias**.
- Once thought of as a regression, not a priori clear the necessity of OU

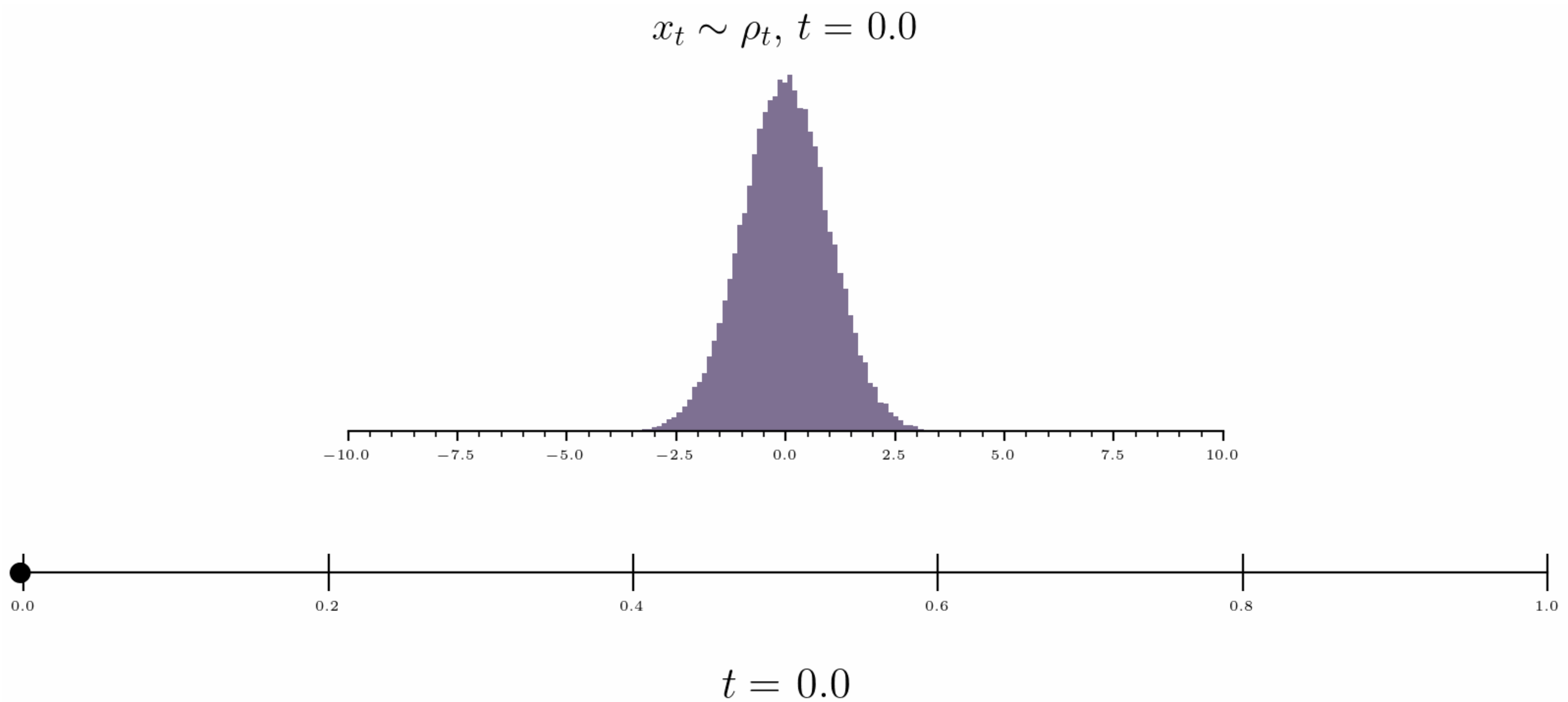
Motivating the Interpolant

How can we work exactly on $t \in [0,1]$ with arbitrary ρ_0 and ρ_1, build a connection between them, and get the velocity $b(t, x)$ directly?

Stochastic Interpolants

Interpolant Function $I(t, x_0, x_1)$

- A function of x_0, x_1, and time t with b.c.'s: $I_{t=0} = x_0$ and $I_{t=1} = x_1$
- Example: $I(t, x_0, x_1) = (1-t)x_0 + tx_1$



If x_0, x_1 drawn independently, then $I(t, x_0, x_1)$ is a **stochastic process which samples** $x_t \sim \rho(t, x)$

$$\rho(t, x) = \mathbb{E}_{\rho_0,\rho_1}\left[\delta\big(x - I(t, x_0, x_1)\big)\right]$$

Interpolant Density

Stochastic Interpolant: $I(t, x_0, x_1)$

$\rho(t, x)$ satisfies continuity equation

$$\rho(t, x) = \mathbb{E}_{\rho_0,\rho_1}\left[\delta\big(x - I(t, x_0, x_1)\big)\right]$$

$$\partial_t \rho(t, x) + \nabla \cdot \big(b(t, x)\rho(t, x)\big) = 0$$

Why? Chain rule gives the current density

$$\partial_t \rho(t, x) = -\mathbb{E}_{\rho_0,\rho_1}[\partial_t I_t \nabla \delta(x - I_t)] \equiv -\nabla \cdot j_t(x)$$

with $j(t, x) = \mathbb{E}_{\rho_0,\rho_1}[\partial_t I_t\, \delta(x - I_t)]$

Stochastic Interpolant: $I(t, x_0, x_1)$

$\rho(t,x)$ satisfies continuity equation

$$\rho(t,x) = \mathbb{E}_{\rho_0,\rho_1}\left[\delta\big(x - I(t,x_0,x_1)\big)\right]$$

$$\partial_t\rho(t,x) + \nabla\cdot\Big(b(t,x)\rho(t,x)\Big) = 0$$

Why? Chain rule gives the current density

$$\partial_t\rho(t,x) = -\,\mathbb{E}_{\rho_0,\rho_1}[\partial_t I_t \nabla\delta(x - I_t)] \equiv -\nabla\cdot j_t(x)$$

with $j(t,x) = \mathbb{E}_{\rho_0,\rho_1}[\partial_t I_t\,\delta(x - I_t)]$

$j(t,x)$ allows us to directly write down a velocity field

$b(t,x) = j(t,x)/\rho(t,x)$ if $\rho(t,x) > 0$

Definition of the Interpolant Velocity: $b(t, x)$

Definition: The conditional expectation of a function f of (t, x_0, x_1) given $x_t = x$ is such that

$$\int \mathbb{E}\big[f(t, x_0, x_1) \mid x_t = x\big]\rho(t, x)\, dx = \mathbb{E}_{\rho_0,\rho_1}\big[f(t, x_0, x_1)\big]$$

Definition of the Interpolant Velocity: $b(t, x)$

Definition: The conditional expectation of a function f of (t, x_0, x_1) given $x_t = x$ is such that

$$\int \mathbb{E}\big[f(t, x_0, x_1) \mid x_t = x\big]\rho(t, x)\, dx = \mathbb{E}_{\rho_0,\rho_1}\big[f(t, x_0, x_1)\big]$$

This gives a **simple form for the velocity field** $b(t, x)$

$$b(t, x) = j(t, x)/\rho(t, x)$$

$$= \frac{\mathbb{E}\big[\partial_t I_t \delta\big(x - I_t\big)\big]}{\mathbb{E}\big[\delta\big(x - I_t\big)\big]} \qquad \text{(if } \rho(t, x) \neq 0 \text{ else } 0\text{)}$$

$$= \mathbb{E}\big[\partial_t I(t, x_0, x_1) \,|\, x_t = x\big]$$

$b(t, x)$ *is readily amenable to estimation via evaluations of* I_t *under* ρ_0, ρ_1

Quadratic Loss over $b(t, x)$

__MSA__ & Vanden-Eijnden arXiv:2209.15571 (2022);
Liu et al. arXiv:2209.03003 (2022);
Lipman et al. arXiv:2210.02747 (2022)

Proposition:

The PDF $\rho(t, x)$ satisfying the continuity equation has a velocity field $b(t, x)$ which is the minimizer of a simple quadratic objective

$$L[\hat{b}] = \min_{\hat{b}(t,x)} \int_0^1 \mathbb{E}\left[\, |\, \hat{b}(t, x_t) - \partial_t I(t, x_0, x_1) \,|^2 \right] dt$$

$$= \left(\, |\, \hat{b}(t, x_t) \,|^2 - 2\partial_t I(t, x_0, x_1) \cdot \hat{b}(t, x_t) \right) dt + \text{const}$$

where $x_t = I(t, x_0, x_1)$.

Quadratic Loss over $b(t,x)$

MSA *& Vanden-Eijnden arXiv:2209.15571 (2022);*
Liu et al. arXiv:2209.03003 (2022);
Lipman et al. arXiv:2210.02747 (2022)

Proposition:

The PDF $\rho(t,x)$ satisfying the continuity equation has a velocity field $b(t,x)$ which is the minimizer of a simple quadratic objective

$$L[\hat{b}] = \min_{\hat{b}(t,x)} \int_0^1 \mathbb{E}\left[|\hat{b}(t,x_t) - \partial_t I(t,x_0,x_1)|^2 \right] dt$$
$$= \left(|\hat{b}(t,x_t)|^2 - 2\partial_t I(t,x_0,x_1)\cdot \hat{b}(t,x_t) \right) dt + \text{const}$$

where $x_t = I(t,x_0,x_1)$.

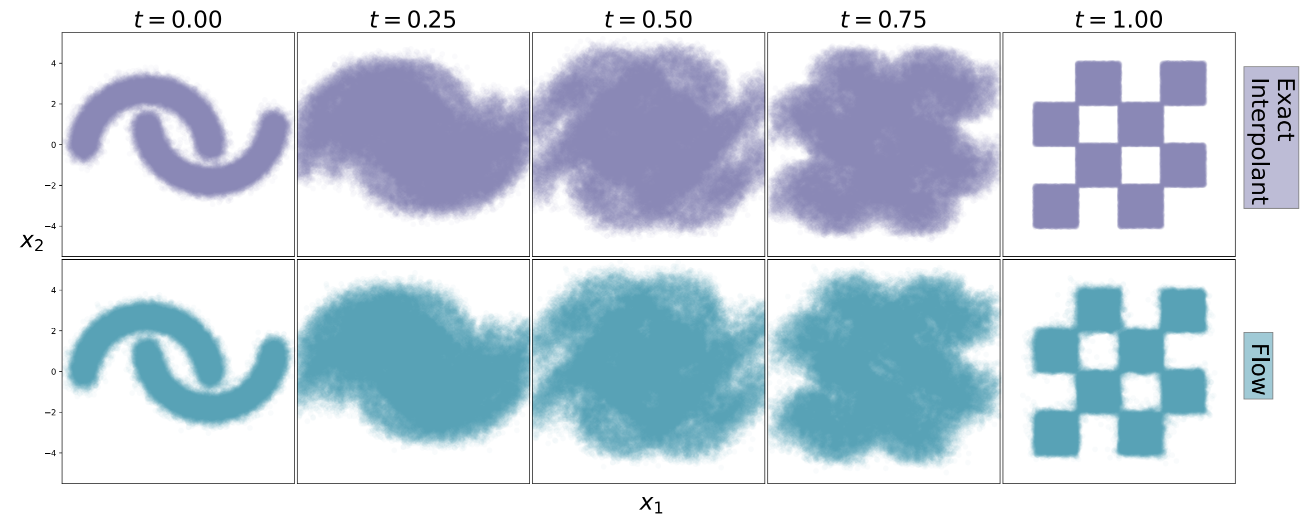
- Loss is directly estimable over ρ_0, ρ_1
- Generative model connects *any* two densities, does not require OU process
- Likelihood and sampling available via fast ODE integrators
- Loss bounds Wasserstein-2 between $\rho(1,x)$ and ρ_1 (Gronwall)

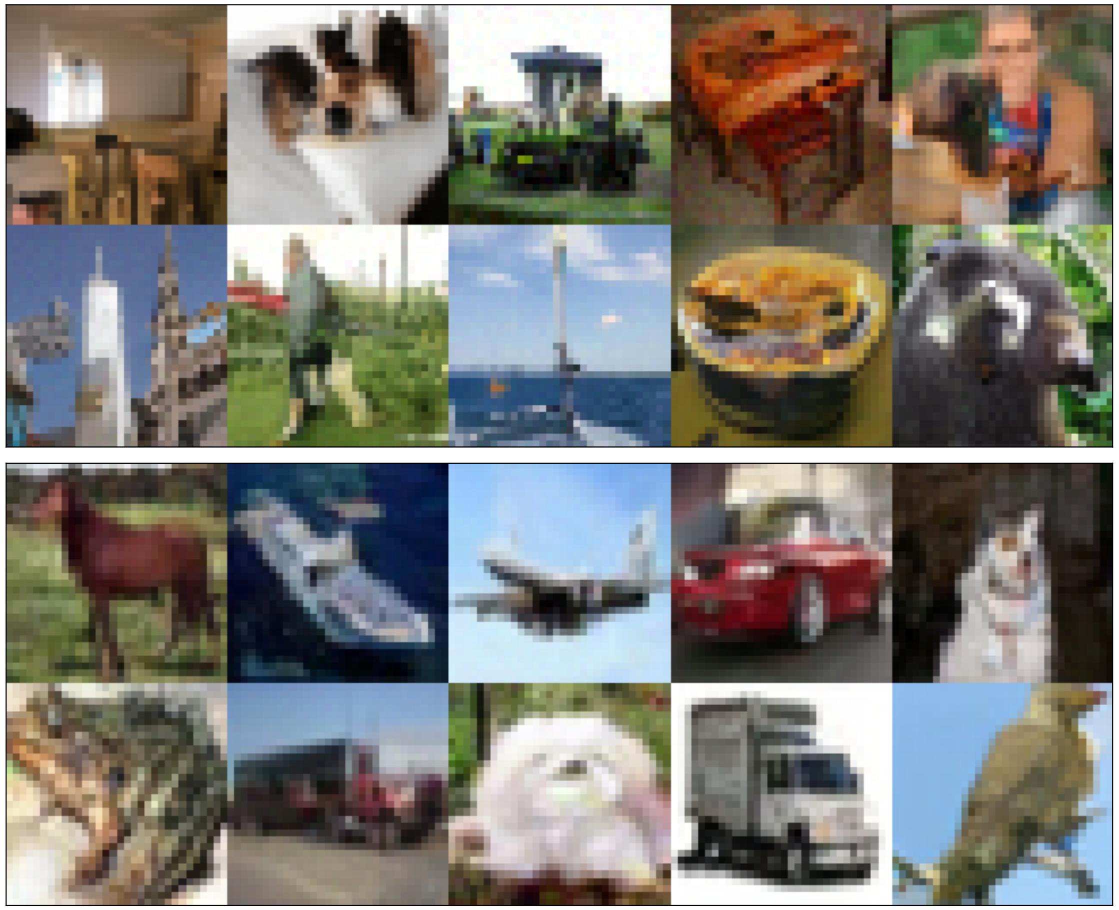
Numerical examples

Toy non-Gaussian densities

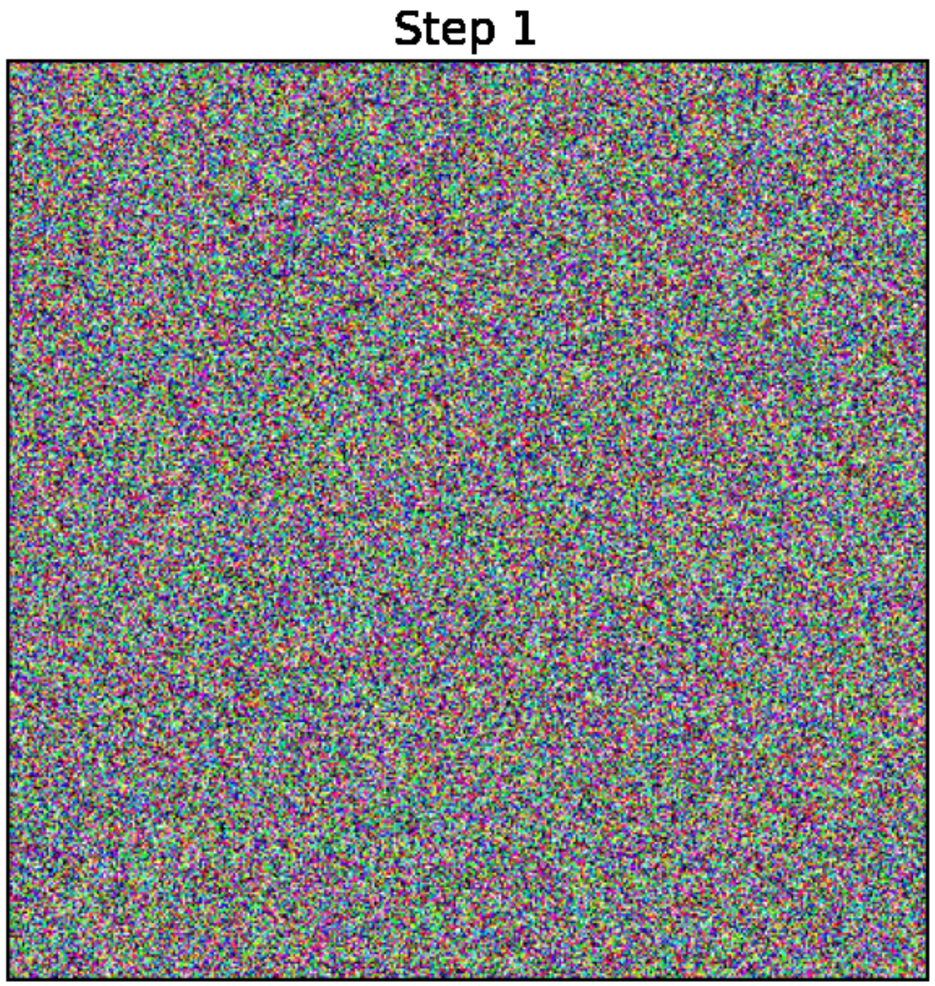
&

Image generation



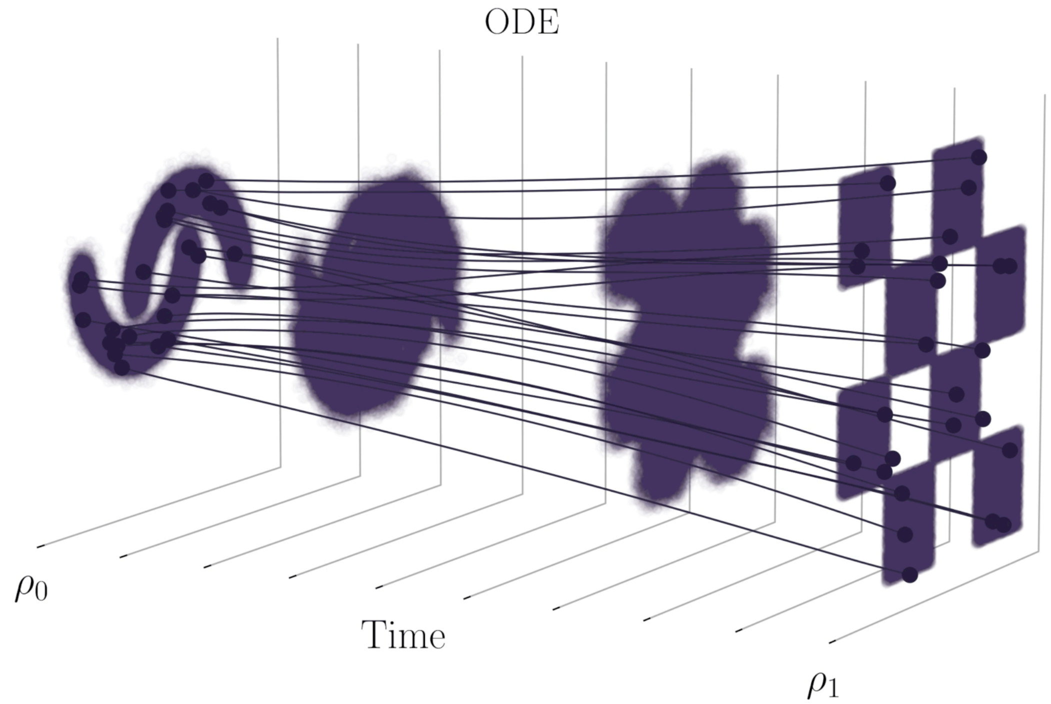


Cifar-10, image-net

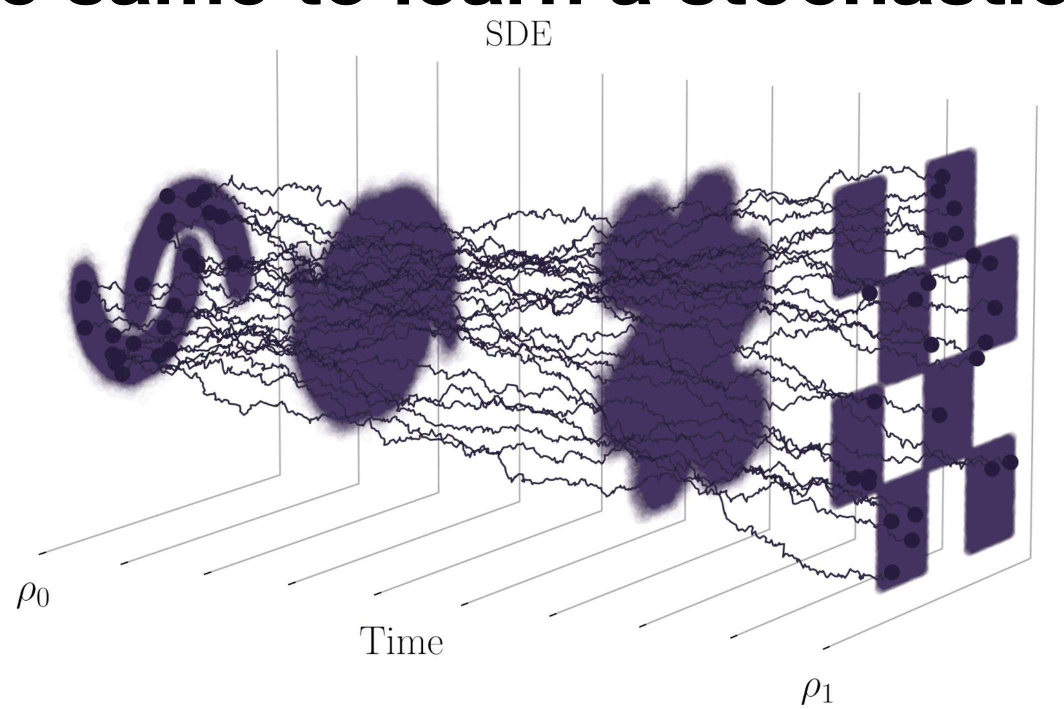


128×128 res flowers

The interpolant paradigm gave us a deterministic flow map between arbitrary densities ρ_0 and ρ_1



Can we do the same to learn a stochastic dynamics?



The interpolant score $s(t, x)$

__MSA__, Boffi, Vanden-Eijnden arXiv:2303.08797 (2023)

Introduce Gaussianity into the interpolant

$$x_t = I(t, x_0, x_1) + \gamma(t)z$$

where $z \sim \mathrm{N}(0,1)$
and $\gamma(0) = \gamma(1) = 0$
e.g. $\gamma(t) = \sqrt{t(1-t)}$

Proposition:
$\rho(t, x)$ satisfies a transport equation as before, with $b(t, x)$ of the form

$$b(t, x) = \mathbb{E}\left[\partial_t I(t, x_0, x_1) + \partial_t \gamma(t) z \,|\, x_t = x\right]$$

Moreover, the score of $\rho(t, x)$ is given by

$$s(t, x) = -\gamma(t)^{-1}\mathbb{E}\left[z \,|\, x_t = x\right]$$

which minimizes

$$L[\hat{s}] = \int \mathbb{E}\left[\frac{1}{2}|\hat{s}(t, x_t)|^2 + \gamma(t)^{-1} z \cdot \hat{s}(t, x_t)\right] dt$$

Unifying Flows and Diffusions

MSA, *Boffi, Vanden-Eijnden arXiv:2303.08797 (2023);*

The score allows us to, like in the case of diffusion models, define a generative stochastic dynamics, now with tunable diffusivity ϵ

Transport equation

$$\partial_t \rho + \nabla \cdot (b\rho) = 0$$

ODE

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

Just learn $\hat{b}$

Fokker-Planck Equations

$$\partial_t \rho + \nabla \cdot (b^{\mathrm{F/B}}\rho) = \epsilon \Delta \rho$$

where $b^{\mathrm{F/B}} = b \pm \epsilon s$

SDE

$$dX_t^{\mathrm{F/B}} = b_{\mathrm{F/B}}\left(t, X_t^{\mathrm{F}}\right) dt + \sqrt{2\epsilon}\, dW_t^{\mathrm{F/B}}$$

Learn $\hat{b}$ and $\hat{s}$

What is the tradeoff between the two? Is there an ϵ^*?

Bounding the KL between ρ and $\hat{\rho}$

Find Nick!

If $\hat{\rho}$ the density pushed by *estimated* deterministic dynamics $\hat{b}$, then

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0$$

$$\mathrm{KL}(\rho(1) \| \hat{\rho}(1)) = \int_0^1 \int_{\mathbb{R}^d} (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b} - b) \rho \, dx \, dt$$

matching b's does not bound KL, Fisher is uncontrolled by small error in $\hat{b} - b$

If $\hat{\rho}$ the density pushed by *estimated* stochastic dynamics $\hat{b}_{\mathrm{F}} = \hat{b} + \epsilon s$, then

$$\partial_t \hat{\rho} + \nabla \cdot (b^{\mathrm{F}} \hat{\rho}) = \epsilon \Delta \hat{\rho}$$

$$\mathrm{KL}(\rho(1) \| \hat{\rho}(1)) \leq \frac{1}{4\epsilon} \int_0^1 \int_{\mathbb{R}^d} \left| \hat{b}_{\mathrm{F}} - b_{\mathrm{F}} \right|^2 \rho \, dx dt$$

$\hat{b}_{\mathrm{F}} - b_{\mathrm{F}}$ does control KL divergence

What does this mean practically?

ODE vs SDE, numerical experiments

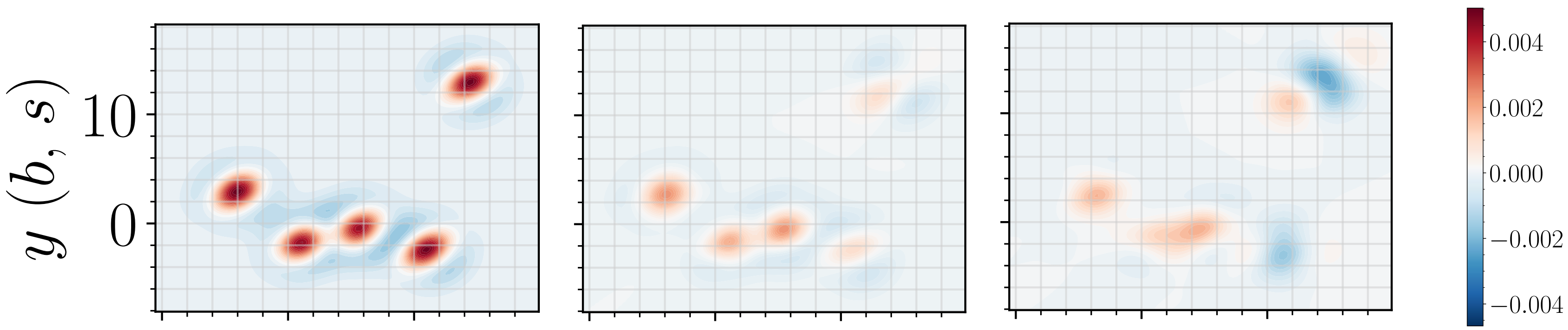
What does this mean practically?

Theory says

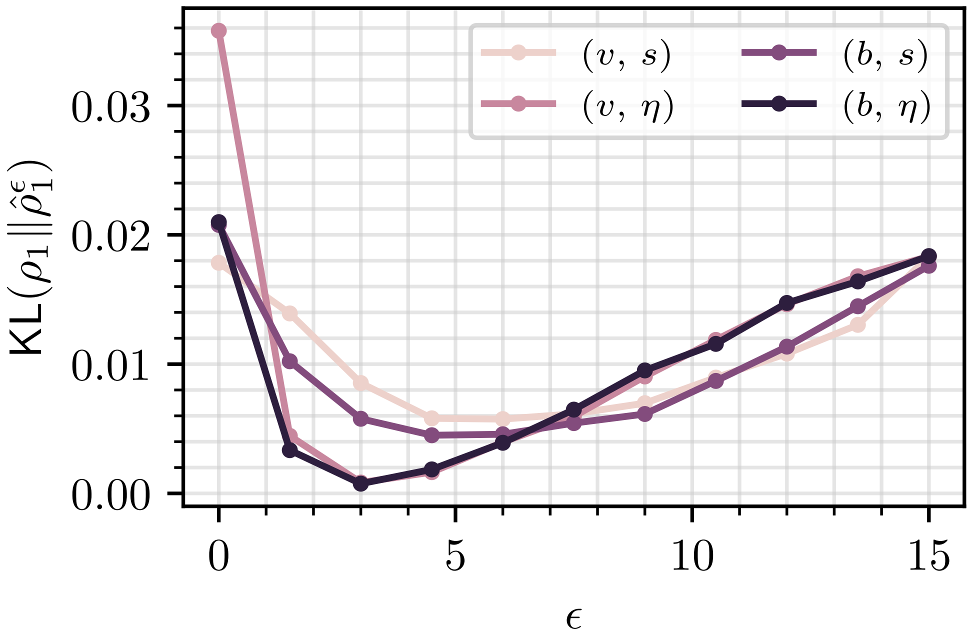
$$\epsilon^* = \left(\frac{L_b[\hat{b}] - \min_{\hat{b}} L_b[\hat{b}]}{L_s[\hat{s}] - \min_{\hat{s}} L_s[\hat{s}]} \right)^{1/2}$$

128 dimensional Gaussian Mixtures

$\hat{\rho}_1(x,y) - \rho_1(x,y)$: **Error in kernel density estimate of 2D cross section**



KL for learned $\hat{b}, \hat{s}$ minimal around $\epsilon \approx 5.0$, then increases



*SDE dominance not necessarily generalize to images

Designing different interpolants: Mirror

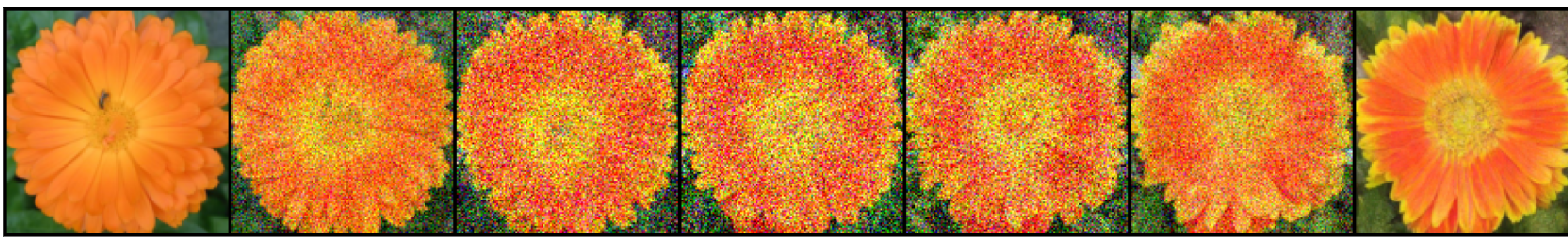
One is free to construct a variety of interpolants, following the rules!

1. Boundary conditions are met
2. The score $s(t, x)$ is available when: $\begin{cases} x_t \text{ includes } \gamma(t)z \\ \text{either } \rho_0 \text{ or } \rho_1 \ \mathrm{N}(0,1) \end{cases}$

Examples! define x_t, state $b(t, x)$

a) *Mirror interpolant*: $x_t = x_1 + \gamma(t)z$

$$b(t, x) = \dot{\gamma}(t)\mathbb{E}\left[z \,|\, x_t = x\right]$$
$$= -\gamma(t)\dot{\gamma}(t)s(t, x)$$



Learn map from ρ_1 — the data density — back to itself!*

(takes *many* SDE steps)

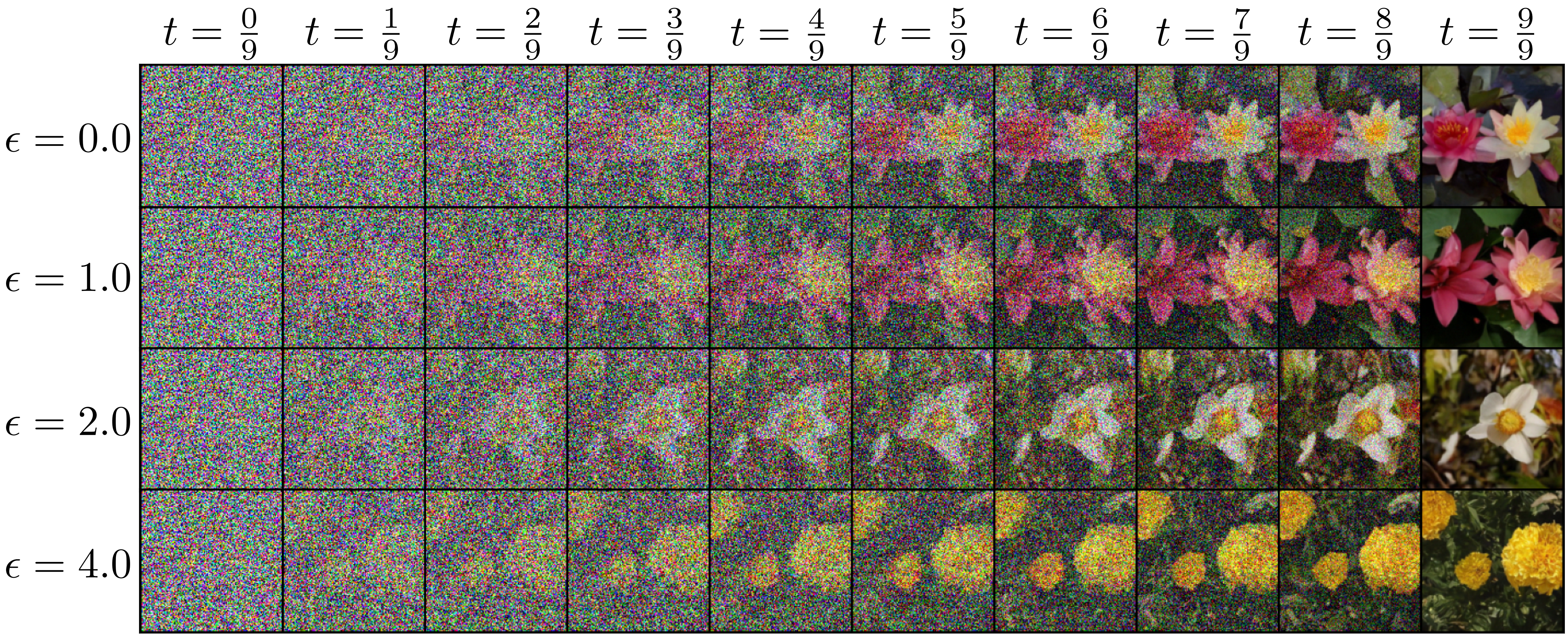
Designing different interpolants: One-sided

b) *One-sided interpolant*: $x_t = \alpha(t)\, z + \beta(t)\, x_1$

$$b(t,x) = \mathbb{E}\left[\dot\alpha(t) z + \dot\beta(t) x_1 \,|\, x_t = x\right]$$

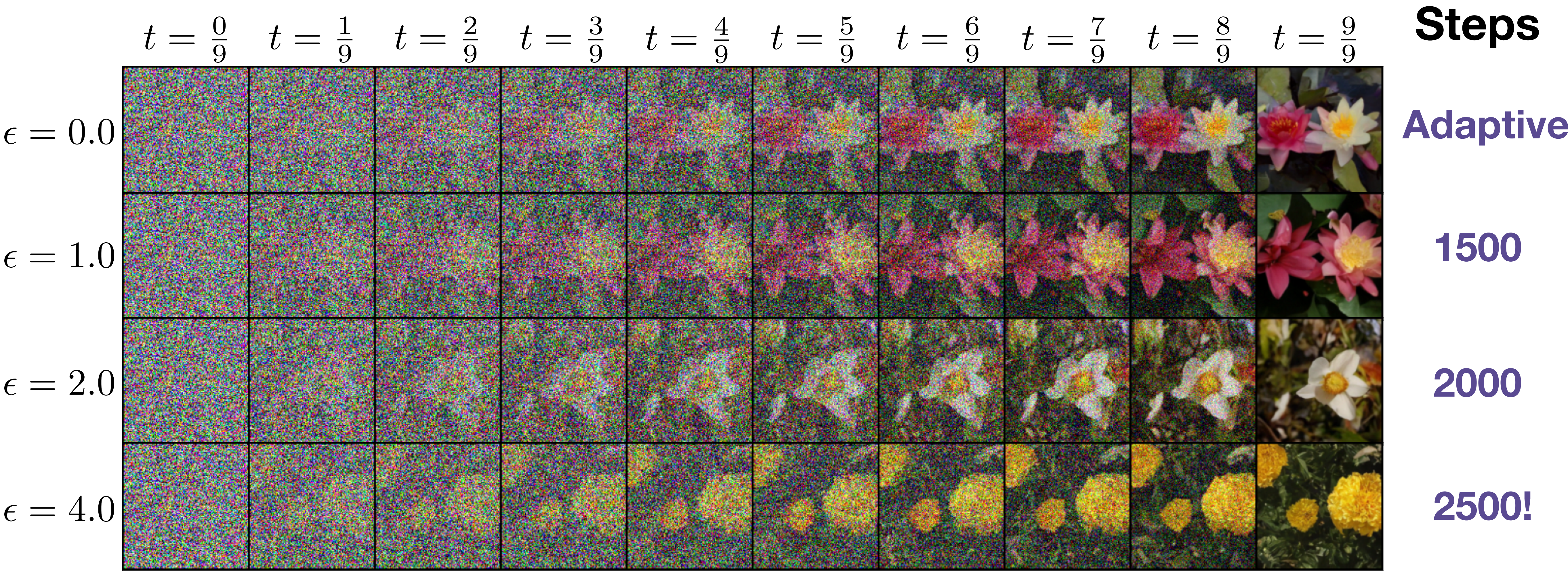
$$s(t,x) = -\,\alpha^{-1}(t)\mathbb{E}\left[z \,|\, x_t = x\right]$$

Demonstrating tunable diffusion



Designing different interpolants: One-sided

Numerical Tradeoffs: Integrating SDEs is harder!



Do slight accuracy benefits seen in GMMs outweigh the numerical costs?

We discussed a method built on stochastic interpolants that:

- allows one to build deterministic **or** stochastic generative models between arbitrary ρ_0, ρ_1
- provides a language for designing new types of maps

Questions going forward:

- Can we use interpolants to study the inductive bias in transport based generative modeling?
- What are realistic assumptions for comparing the ODE and SDE?
  Chen et al, 2305.11798, 2303.03384 Wibisono, 2211.01512
- Better SDE integrators than Euler-Maruyama?
- Applications to variational inference (access to the target $\log \rho_1$)?

Summary and Outlook

We discussed a method built on stochastic interpolants that:

- allows one to build deterministic **or** stochastic generative models between arbitrary ρ_0, ρ_1
- provides a language for designing new types of maps

See the papers (and Nick, here this week!) for:

- Optimizing the transport costs in both the ODE (OT) and SDE (Schrodinger Bridge) setting
- More experimental details
- Some preliminary available code: https://github.com/malbergo/stochastic-interpolants

Thanks !

Backup Slides

Expanded minimizer of MLE object for continuous flows

$$\min_{v} KL(\rho_1 \,||\, \rho(1)) = \min \mathbb{E}_{\rho_1}\left[\log \frac{\rho_1(x)}{\rho(1,x)}\right]$$

$$= \min - \mathbb{E}_{\rho_1}\left[\log \rho(1,x)\right] + C$$

Under reverse dynamics $X_{t=1} = x \sim \rho_1$

$$\min_{v} \mathbb{E}_{\rho_1}\left[\int_0^1 \nabla \cdot b(t, \bar{X}_t(x))dt - \log \rho_0(\bar{X}_{t=0}(x))\right] \qquad \text{s.t.} \quad \dot{\bar{X}}_t(x) = b(t, X_t(x))$$

Proof of score $s(t, x)$ minimizer (1/2)

Let $g(t, k) = \mathbb{E}\left[e^{ikx_t}\right] = \mathbb{E}\left[e^{ikI_t}\right] E\left[e^{ikz}\right]$ be the characteristic function of $\rho(t, x)$

Then $g(t, k) = g_0(t, k) e^{-\frac{1}{2}\gamma^2(t)|k|^2}$

Consider the term $\mathbb{E}\left(ze^{i\gamma(t)kz}\right)$ for which:

$$\mathbb{E}\left[ze^{i\gamma(t)k\cdot z}\right] = -\gamma^{-1}(t)\left(i\partial_k\right)\mathbb{E}e^{i\gamma(t)k\cdot z} = i\gamma(t)ke^{-\frac{1}{2}\gamma^2(t)|k|^2}$$

Then:

$$\mathbb{E}\left[ze^{ikx_t}\right] = \mathbb{E}\left[ze^{ik\gamma(t)z}\right]\mathbb{E}\left[e^{ikI_t}\right] = i\gamma(t)kg(t, k)$$

Relatedly using the conditional expectation:

$$\mathbb{E}\left[ze^{ikx_t}\right] = \int \mathbb{E}\left[ze^{ikx_t} \mid x_t = x\right]\rho(t, x)dx = \int \mathbb{E}\left[z \mid x_t = x\right]e^{ikx}\rho(t, x)dx = F\left(\mathbb{E}\left[z \mid x_t = x\right]\rho(t, x)\right)$$

Where F is the Fourier operator

Proof of score $s(t, x)$ minimizer (1/2)

From before, note that the RHS of :

$$\mathbb{E}\left[ze^{ikx_t}\right] = i\gamma(t)kg(t,k)$$

is the Fourier transform $F(-\gamma(t)\nabla\rho(t,x))$.

So $\mathbb{E}\left[ze^{ikx_t}\right] = F(\mathbb{E}\left[z\,|\,x_t = x\right]) = F(-\gamma(t)\nabla\rho(t,x))$

Using $\nabla\rho/\rho = \nabla\log\rho$, we have

$$-\gamma(t)^{-1}\mathbb{E}\left[z\,|\,x_t = x\right] = \nabla\log\rho(t,x)$$

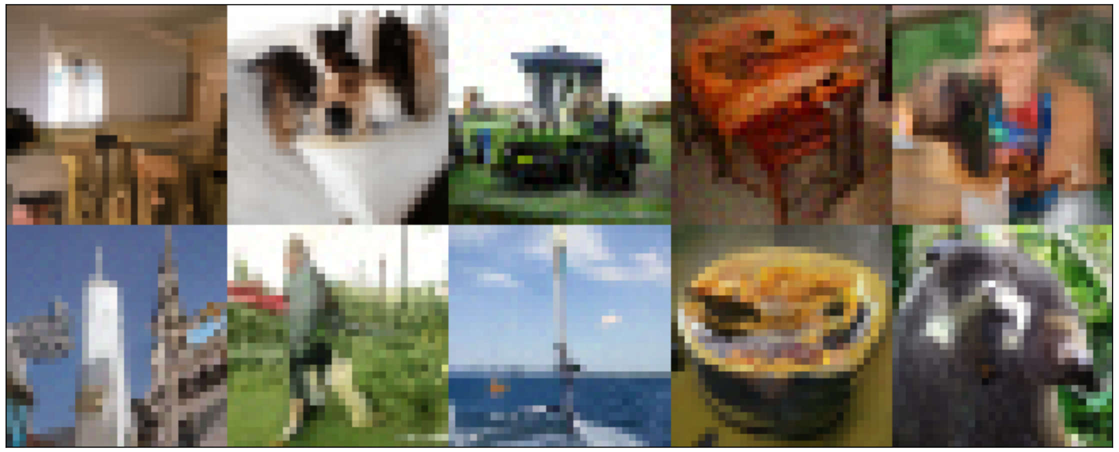
ODE vs SDE, numerical experiments

What does this mean practically?

Theory says

$$\epsilon^* = \left(\frac{L_b[\hat{b}] - \min_{\hat{b}} L_b[\hat{b}]}{L_s[\hat{s}] - \min_{\hat{s}} L_s[\hat{s}]} \right)^{1/2}$$

Image experiments



	FID
DDPM [20]	6.99
ScoreFlow [45]	5.68
Flow matching OT [31]	5.02
Interpolant	
$\varepsilon = 0.0$	6.28
$\varepsilon = 0.5$	6.95
$\varepsilon = 1.0$	6.45

Frechet Inception Distance (FID)

SBDM Interpolant

$$y_t = x_0 e^{-t} + \sqrt{1 - e^{-2t}} z, \quad x_0 \sim \rho_0, \quad z \sim \mathrm{N}(0, Id), \quad t \in [0, \infty)$$