ICML 2022

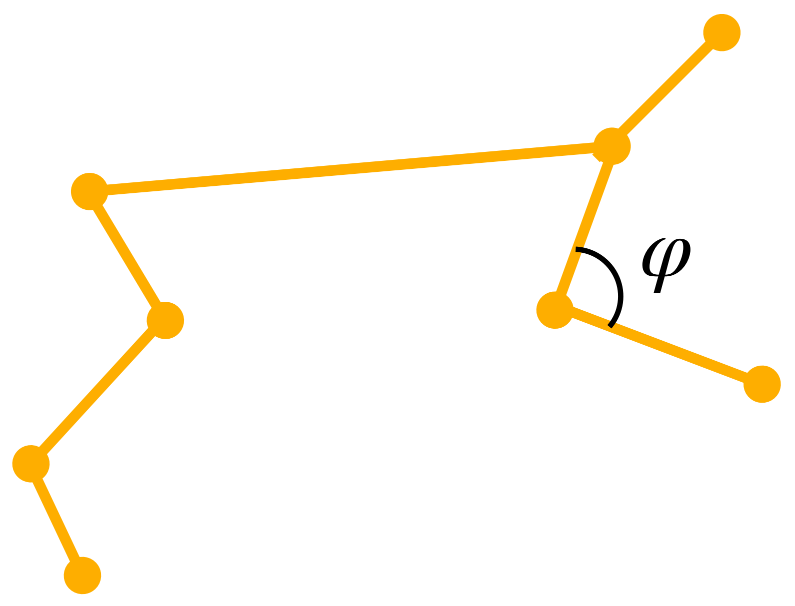
Matching Normalizing Flows and Probability Paths on Manifolds

Heli Ben-Hamu*, Samuel Cohen*, Joey Bose, Brandon Amos, Aditya Grover,
Maximilian Nickel, Ricky T. Q. Chen, Yaron Lipman

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

FACEBOOK AI

Generative Modeling on Manifolds

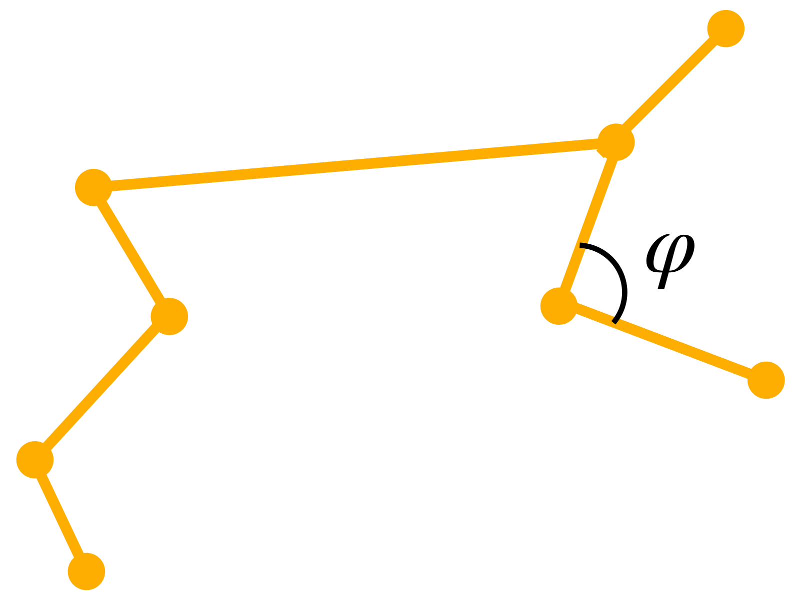


Fire locations on earth [Mathieu et al. ’20]

Generative Modeling on Manifolds

Motivation

Scientific data from various fields lie on manifolds

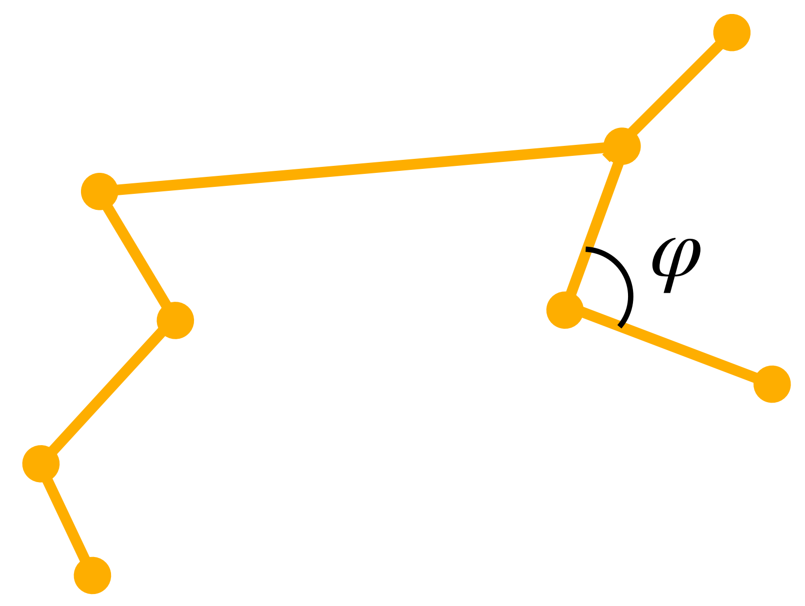


Fire locations on earth [Mathieu et al. ’20]

Generative Modeling on Manifolds

Motivation

Scientific data from various fields lie on manifolds



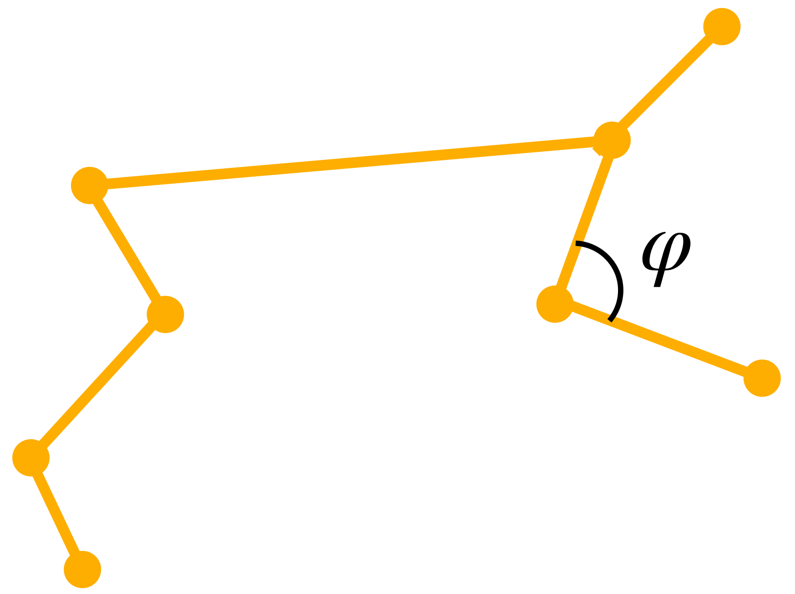
Continuous Normalizing Flows (CNFs) on manifolds

Fire locations on earth [Mathieu et al. ’20]

Generative Modeling on Manifolds

Motivation

Scientific data from various fields lie on manifolds



Fire locations on earth [Mathieu et al. ’20]

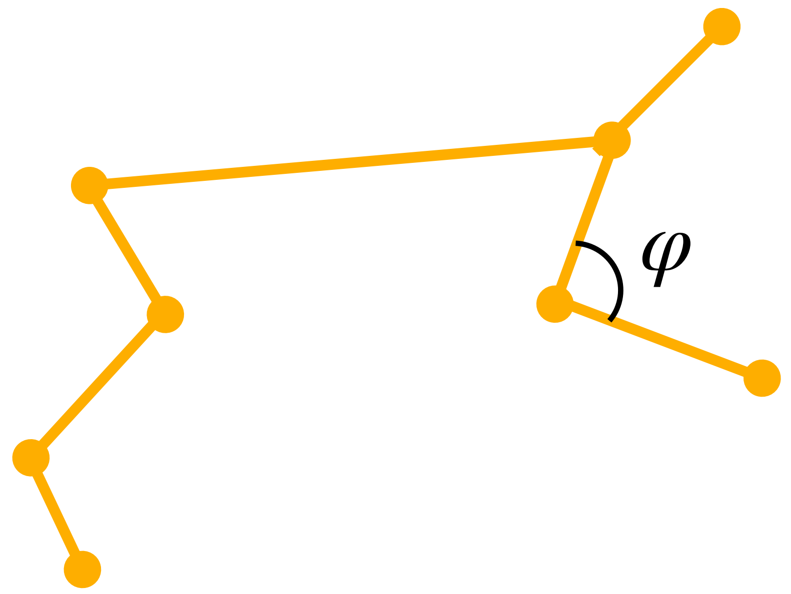
Continuous Normalizing Flows (CNFs) on manifolds

- High computational cost [Mathieu et al. ’20]

Generative Modeling on Manifolds

Motivation

Scientific data from various fields lie on manifolds



Fire locations on earth [Mathieu et al. ’20]

Continuous Normalizing Flows (CNFs) on manifolds

- High computational cost [Mathieu et al. ’20]
- Do not scale to high dimensions [Rozen et al. ’21]

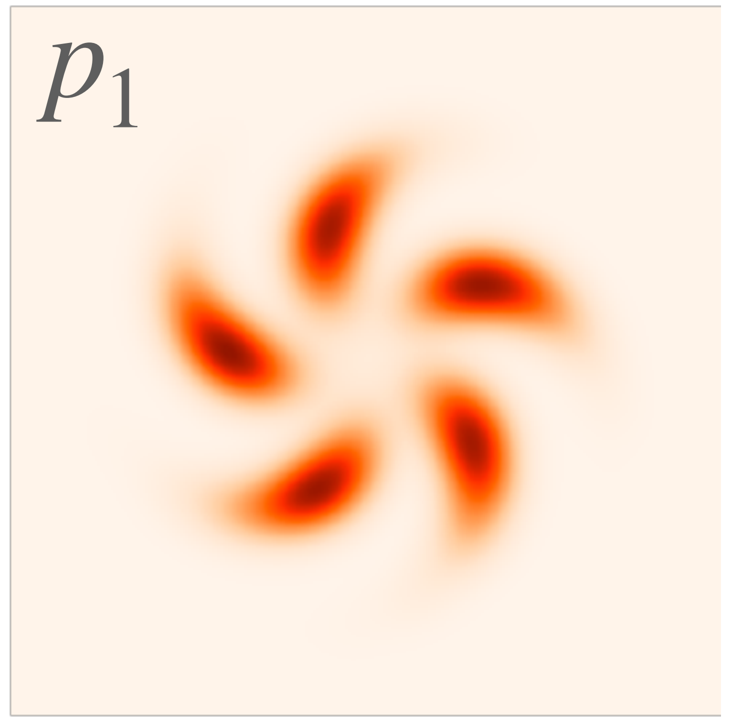
Our Work:

A novel CNF training framework based on

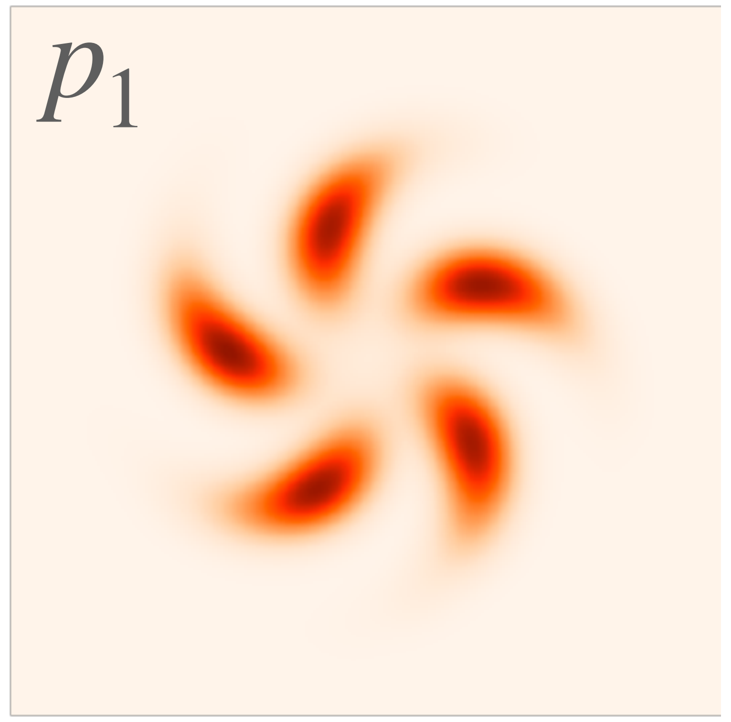
Probability Path Matching

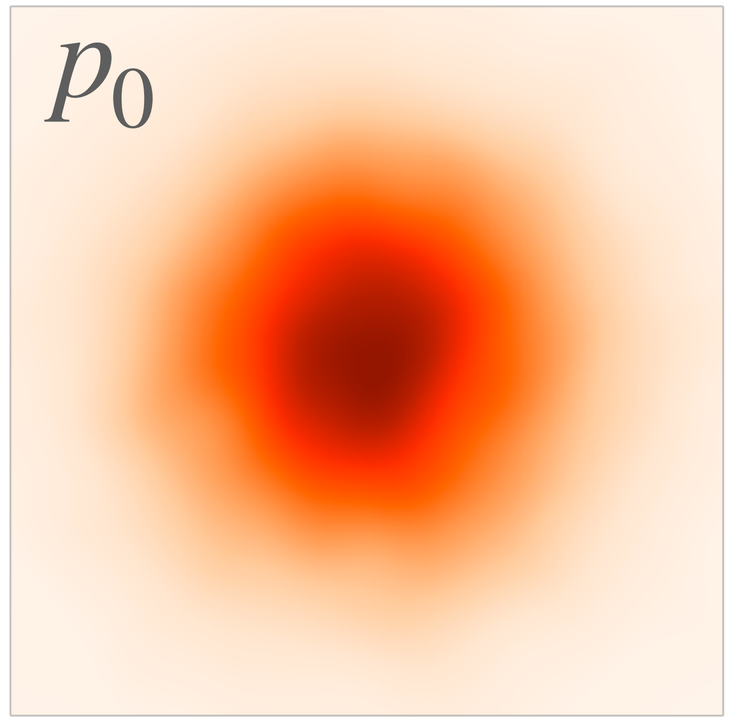
Probability Path Matching

Probability Path Matching

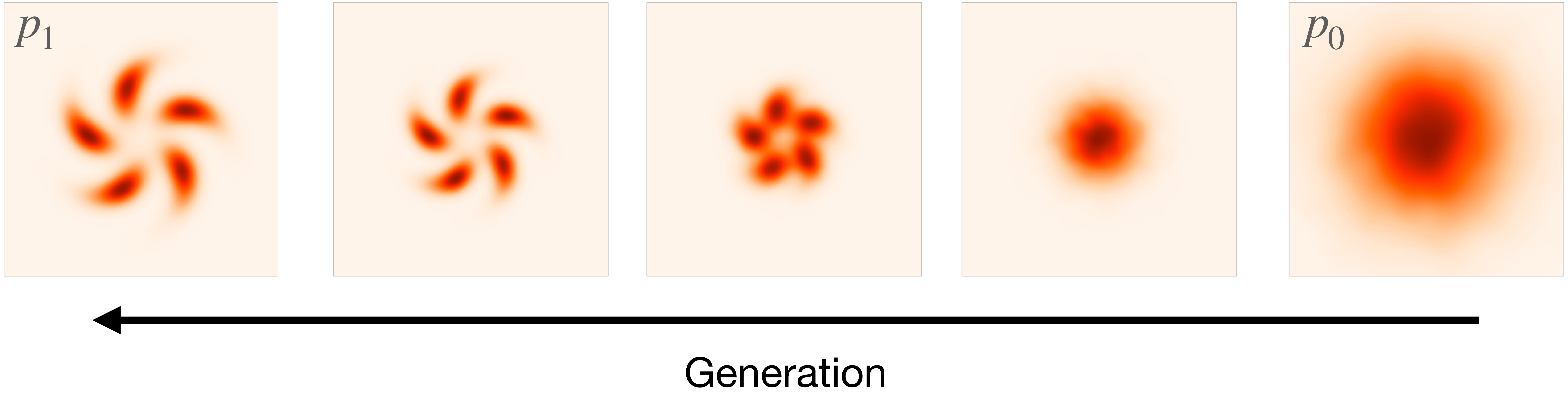


Probability Path Matching

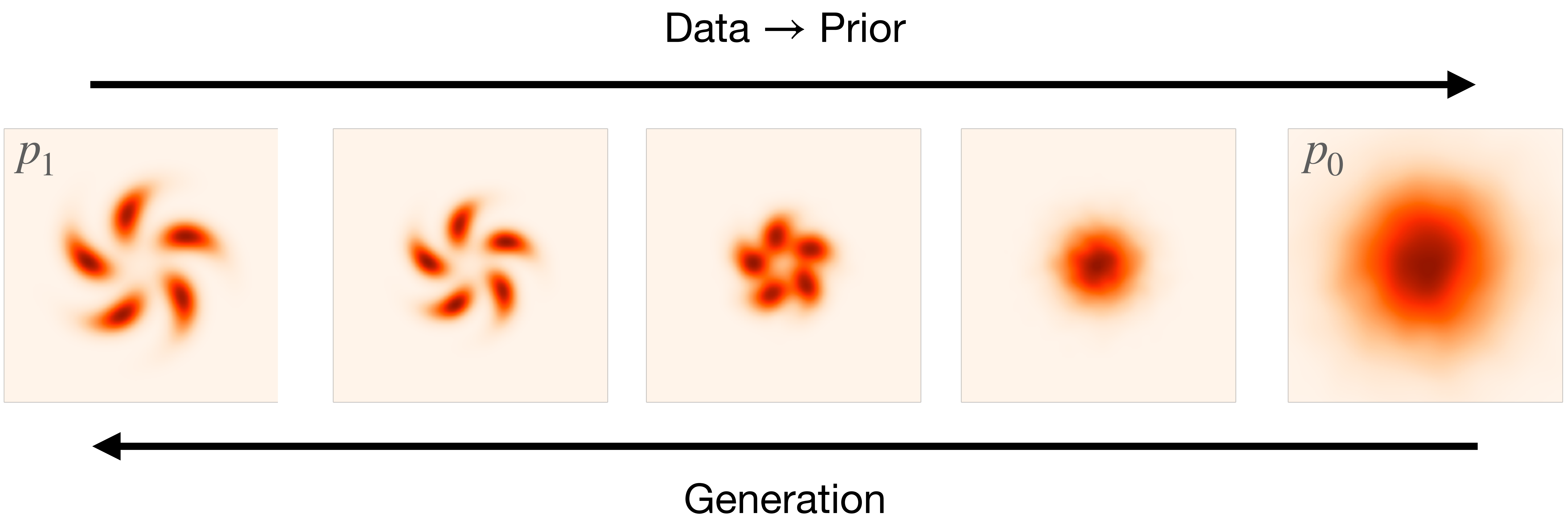




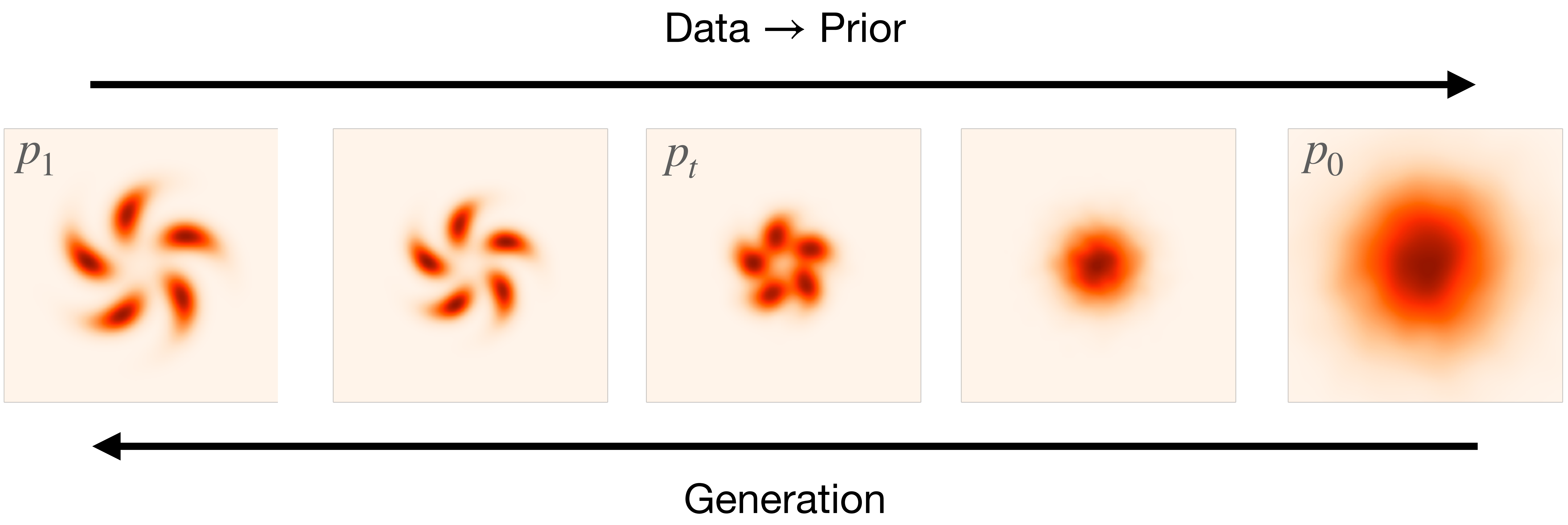
Probability Path Matching



Probability Path Matching

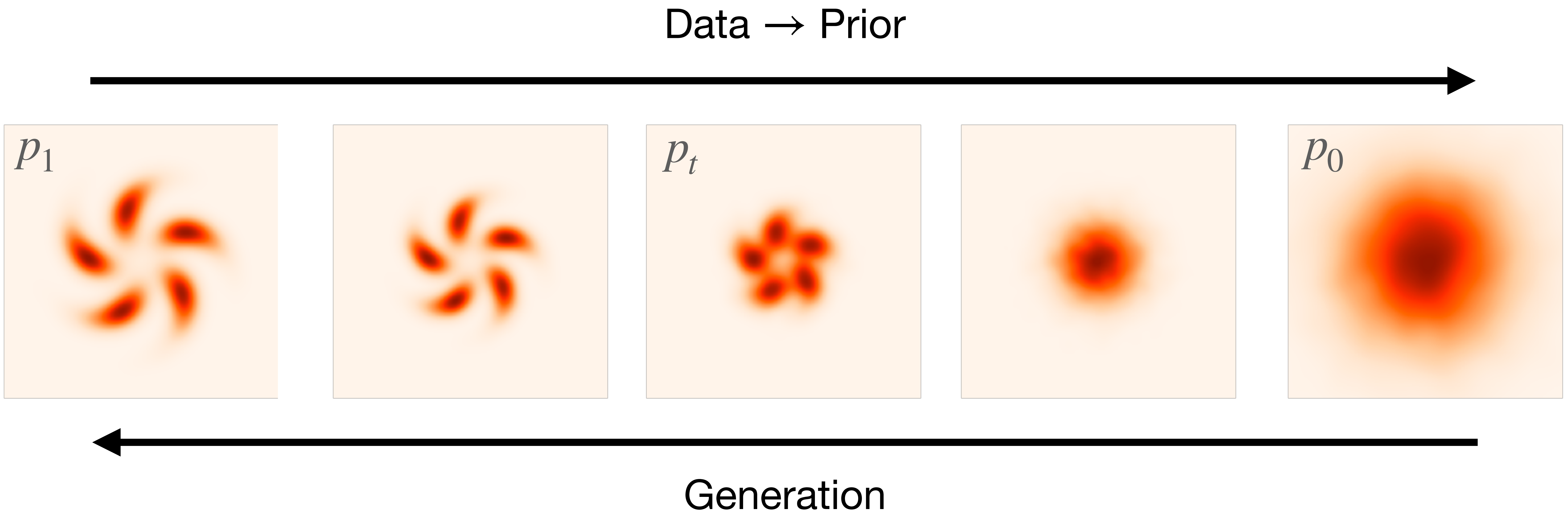


Probability Path Matching



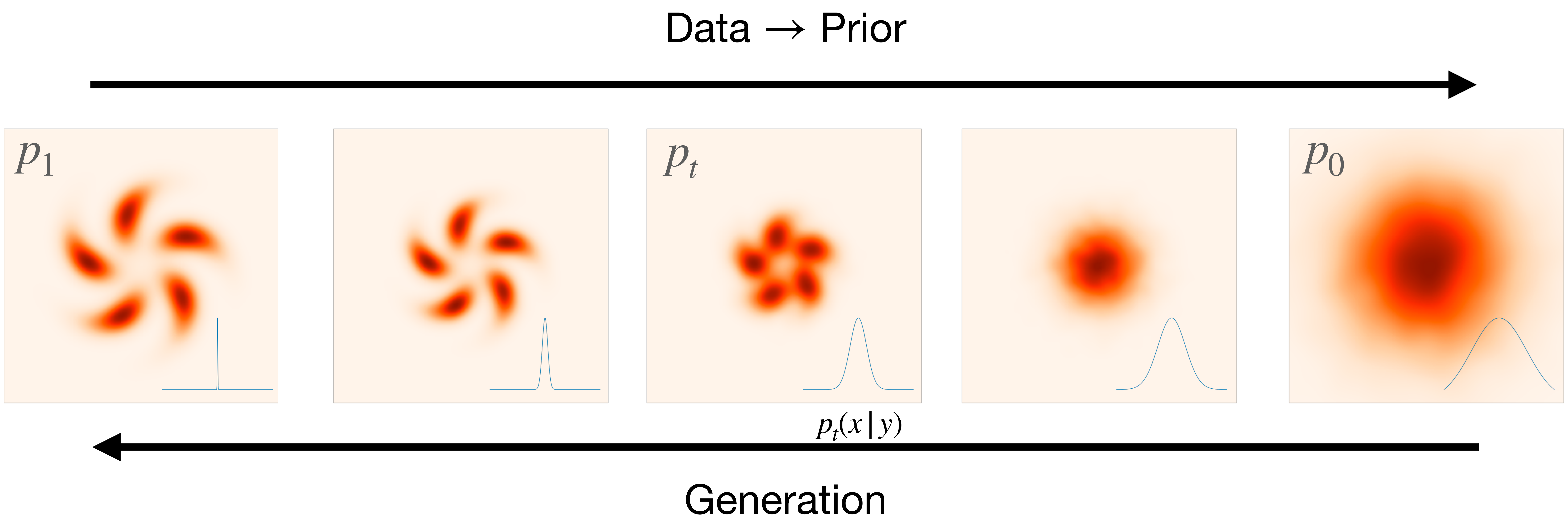
Probability Path - p_t

Probability Path Matching



Path Construction $p_t(x) = \int p_t(x \mid y) p_1(y) dy$

Probability Path Matching



Path Construction $p_t(x) = \int p_t(x\,|\,y)p_1(y)dy$

Probability Path Matching

Target - probability path p_t

Model - probability path q_t with parameters θ

Probability Path Matching

Target - probability path p_t

Model - probability path q_t with parameters θ

Goal: find parameters θ s.t. $q_t \approx p_t$

Probability Path Matching

Target - probability path p_t

Model - probability path q_t with parameters θ

Goal: find parameters θ s.t. $q_t \approx p_t$

$$\min_{\theta} \; \mathrm{d}(p_t \,||\, q_t)$$

Matching CNFs to Target Paths

Target - probability path p_t

Model - probability path q_t with parameters θ

CNF

$$q_t = \phi_{t*} p_0 \qquad \frac{d}{dt}\phi_t = v_\theta(t, \phi_t)$$

$$\min_\theta \ \mathrm{d}(p_t \,||\, q_t)$$

Matching CNFs to Target Paths

Target - probability path p_t

Model - a CNF ϕ_t parametrized by $v_\theta(t, x)$

CNF

$$q_t = \phi_{t*} p_0 \qquad \frac{d}{dt}\phi_t = v_\theta(t, \phi_t)$$

$$\min_\theta \; \mathrm{d}(p_t \,\|\, q_t)$$

s.t. $q_t = \phi_{t*} p_0$

Matching CNFs to Target Paths

Target - probability path p_t

Model - a CNF ϕ_t parametrized by $v_\theta(t, x)$

CNF

$$q_t = \phi_{t*}p_0 \qquad \frac{d}{dt}\phi_t = v_\theta(t, \phi_t)$$

$$\min_\theta \; \mathrm{d}(p_t \,\|\, q_t)$$

s.t. $q_t = \phi_{t*}p_0$

Probability Path Divergence (PPD)

Probability Path Divergence

$$\partial_t p_t + \mathrm{div}(p_t v) = 0$$

Mass Conservation Formula

Probability Path Divergence

$$\partial_t p_t + \mathrm{div}(p_t v) = 0$$

Mass Conservation Formula

$$\Updownarrow$$

$$p_t = \phi_{t*} p_0$$

A CNF parametrized by v

Probability Path Divergence

$$\partial_t \log p_t + \langle \nabla_x \log p_t, v \rangle + \mathrm{div}(v) = 0$$

Logarithmic Mass Conservation Formula

$$\Updownarrow$$

$$p_t = \phi_{t*} p_0$$

Probability Path Divergence

$$\partial_t \log p_t + \langle \nabla_x \log p_t, v \rangle + \operatorname{div}(v) = 0$$

Logarithmic Mass Conservation Formula

$$\Updownarrow$$

$$p_t = \phi_{t*} p_0$$

A CNF parametrized by v

Probability Path Divergence

We define a family of probability path divergences (PPD) with integer parameter $\ell \geq 1$:

$$\mathrm{d}_\ell(p_t \parallel q_t) = \mathbb{E}_{t,x\sim p_t}\left|\partial_t \log p_t + \langle \nabla_x \log p_t, v \rangle + \mathrm{div}(v)\right|^\ell$$

Probability Path Divergence

We define a family of probability path divergences (PPD) with integer parameter $\ell \geq 1$:

$$\mathrm{d}_\ell(p_t \,\|\, q_t) = \mathbb{E}_{t,x\sim p_t} |\partial_t \log p_t + \langle \nabla_x \log p_t, v \rangle + \mathrm{div}(v)|^\ell$$

- $\mathrm{d}_\ell(p_t \,\|\, q_t) \geq 0$ by construction
- $\mathrm{d}_\ell(p_t \,\|\, q_t) = 0$ iff $p_t = q_t,\ \forall t \in [0,1]$

Matching CNFs to Target Paths

Target - probability path p

Model - a CNF ϕ_t parametrized by $v_\theta(t, x)$

$$\min_\theta \boxed{\mathrm{d}(p_t \,\|\, q_t)}$$

s.t. $q_t = \phi_{t*} p_0$

Matching CNFs to Target Paths

Target - probability path p

Model - a CNF ϕ_t parametrized by $v_\theta(t,x)$

$$\min_\theta \; \mathbb{E}_{t,x\sim p_t} \left| \partial_t \log p_t + \langle \nabla_x \log p_t, v_\theta \rangle + \mathrm{div}(v_\theta) \right|^\ell$$

$$\text{s.t.} \quad q_t = \phi_{t*} p_0$$

Matching CNFs to Target Paths

Target - probability path p

Model - a CNF ϕ_t parametrized by $v_\theta(t, x)$

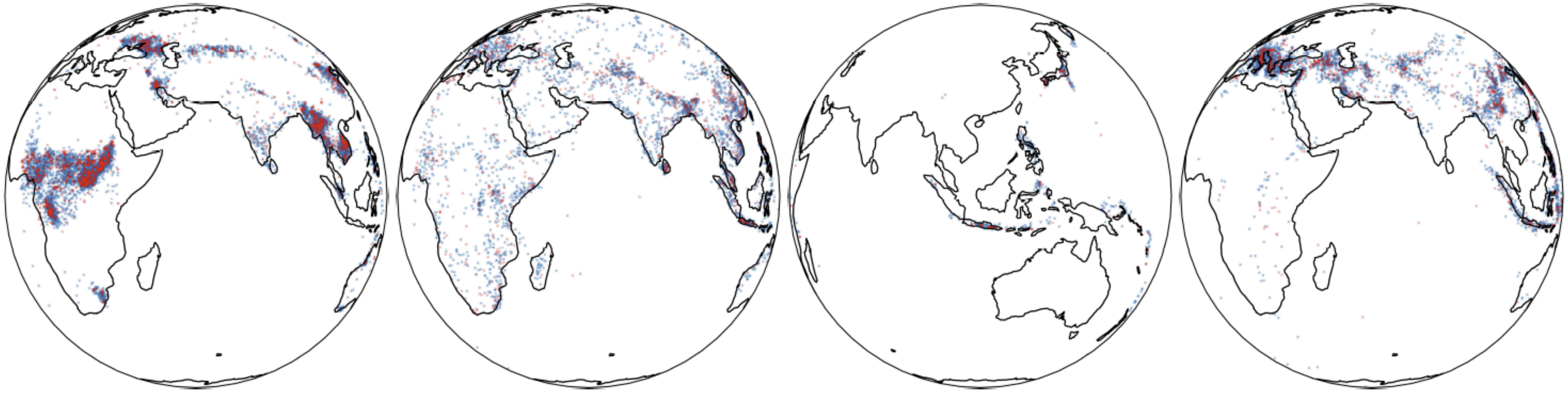
Optimization and evaluation of PPD only require access to v_θ

$$\min_\theta \mathbb{E}_{t,x\sim p_t} \left| \partial_t \log p_t + \langle \nabla_x \log p_t, v_\theta \rangle + \mathrm{div}(v_\theta) \right|^\ell$$

Earth and Climate Dataset

Dataset	Earthquake	Flood	Fire	Volcano
Mixture vMF	$0.59_{\pm 0.01}$	$1.09_{\pm 0.01}$	$-0.23_{\pm 0.02}$	$-0.31_{\pm 0.07}$
Stereographic	$0.43_{\pm 0.04}$	$0.99_{\pm 0.04}$	$-0.40_{\pm 0.06}$	$-0.64_{\pm 0.20}$
Riemannian	$0.19_{\pm 0.04}$	$0.90_{\pm 0.03}$	$-0.66_{\pm 0.05}$	$-0.97_{\pm 0.15}$
Moser Flow	$-0.09_{\pm 0.02}$	$0.62_{\pm 0.04}$	$-1.03_{\pm 0.03}$	$-2.02_{\pm 0.42}$
CNFM	$\mathbf{-0.38}_{\pm 0.01}$	$\mathbf{0.25}_{\pm 0.02}$	$\mathbf{-1.40}_{\pm 0.02}$	$\mathbf{-2.38}_{\pm 0.16}$

NLL scores

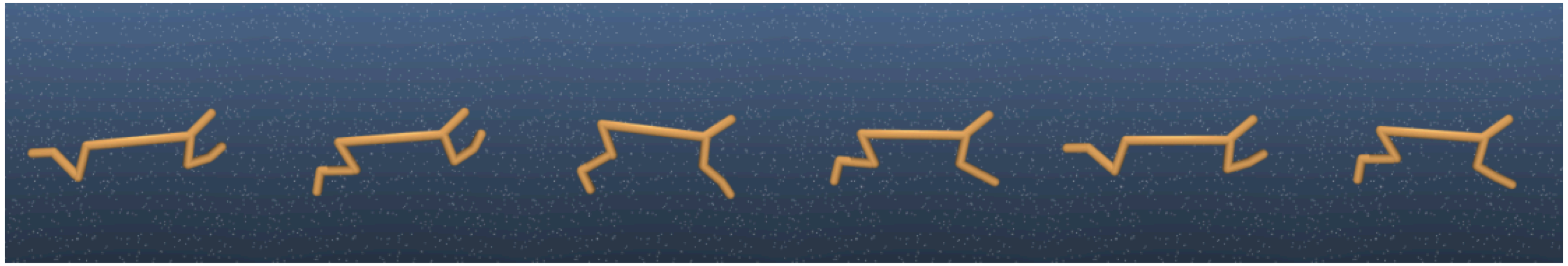


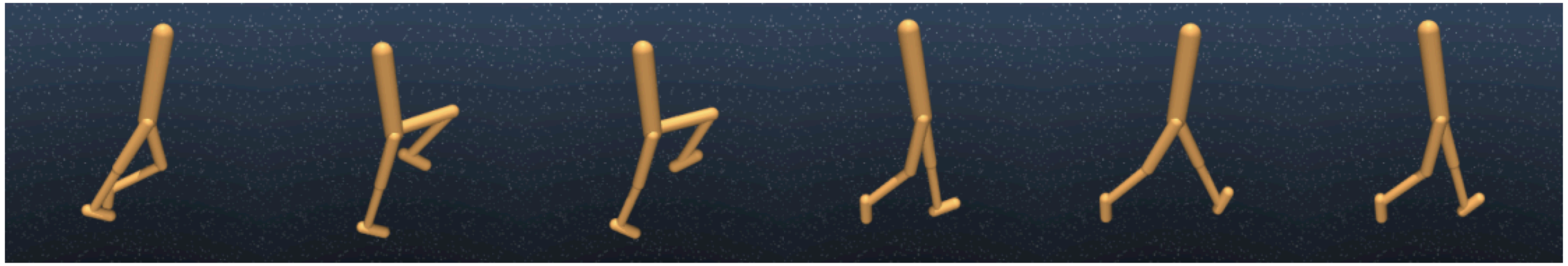
Fire Flood Volcano Quakes

Blue - generated samples, Red - test samples

Product of Manifolds - Robotics

$$\mathscr{M} = \mathbb{R}^3 \times (\mathcal{S}^1)^6$$





$$\mathscr{M} = \mathbb{R}^3 \times (\mathcal{S}^1)^8 \times (\mathcal{S}^3)^6$$

Limitations and Conclusions

Conclusions

- Introduced a **novel** divergence allowing **scalable** training of CNFs
- First application to **higher dimensions** on manifolds (~30)

Limitations

- Scaling to even higher dimensions (>100):
  - Biased loss gradients

Thank You!

Coming Soon!

github.com/helibenhamu/CNFM