Momentum Generative Model

LoGG Presentation

Presenter: Tianrong Chen date:11/06/2023

Agenda

- **Introduction of Dynamical Generative Model**
  - Diffusion Model
  - Flow Matching
  - Applications: Generative Modeling and Trajectory Inference
- **Momentum Generative Model**
  - Acceleration Generative Model (AGM)
  - Momentum Schrödinger Bridge (mSB)
- **Q&A**

Dynamical Generative Modeling

$$dx_t = f(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

$$dx_t = b(x_t, t)dt + g_t dw_t$$

Task:

1. **Generative Modeling:** Sample data from data distribution p_0
2. **Trajectory Inference:** sample data and intermediate trajectories in between.

How: Construct transportation map between prior p_1 and p_0 by Dynamical System: SDE (Diffusion Model [1]), ODE (Flow Matching[2]) or PDE (Neural Fourier Operator[3])

[1] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).
[2] Lipman, Yaron, et al. "Flow matching for generative modeling." *arXiv preprint arXiv:2210.02747* (2022).
[3] Li, Zongyi, et al. "Fourier neural operator for parametric partial differential equations." *arXiv preprint arXiv:2010.08895* (2020).

Dynamical Generative Modeling

$$dx_t = f(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

$$dx_t = b(x_t, t)dt + g_t dw_t$$

Diffusion Model:

Predefine diffusion process.

1. Choose $f(\cdot)$ and $g_t(\cdot)$

Learn the reverse process (Time reversal Theorem)

1. $b(\cdot) := f - g^2 \nabla_x \log p(\cdot)$

Dynamical Generative Modeling

$$dx_t = f(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

$$dx_t = b(x_t, t)dt + g_t dw_t$$

Diffusion Model:

Predefine diffusion process.

1. Choose $f(\cdot)$ and $g_t(\cdot)$

Learn the reverse process (Time reversal Theorem)

1. $b(\cdot) := f - g^2 \nabla_x \log p(\cdot)$

Flow Matching:

Predefine data pairing $(x_0, x_1) \sim p_0 \otimes p_1$

1. Construct linear interpolation vector field $f(\cdot, t \,|\, x_0, x_1) := x_1 - x_0$, $g_t \equiv 0$

Learn the reverse process

1. $b(\cdot) := -f(\cdot)$

Dynamical Generative Modeling

$$dx_t = f(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

$$dx_t = b(x_t, t)dt + g_t dw_t$$

Diffusion Model:

Predefine diffusion process.

1. Choose $f(\cdot)$ and $g_t(\cdot)$

Learn the reverse process (Time reversal Theorem)

1. $b(\cdot) := f - g^2 \nabla_x \log p(\cdot)$

Flow Matching:

Predefine data pairing $(x_0, x_1) \sim p_0 \otimes p_1$

~~1. Construct linear interpolation vector field~~ $\cancel{f(\cdot, t \mid x_0, x_1) := x_1 - x_0, g_t \equiv 0}$

Learn the vector field

$b(\cdot) := -(x_1 - x_0)$

Dynamical Generative Modeling

$$dx_t = f(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

$$dx_t = b(x_t, t)dt + g_t dw_t$$

Connection?

Diffusion Model:

Predefine diffusion process.

1. Choose $f(\cdot)$ and $g_t(\cdot)$

Learn the reverse process

1. $b(\cdot) := f - g^2 \nabla_x \log p(\cdot)$

Flow Matching:

Predefine data pairing $(x_0, x_1) \sim p_0 \otimes p_1$

1. Construct linear interpolation vector field $f(\cdot, t \,|\, x_0, x_1) := x_1 - x_0$, $g_t \equiv 0$

Learn the reverse process

1. $b(\cdot) := -f(\cdot)$

Dynamical Generative Modeling

$$dx_t = f(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

$$dx_t = b(x_t, t)dt + g_t dw_t$$

Diffusion Model:

Predefine diffusion process.

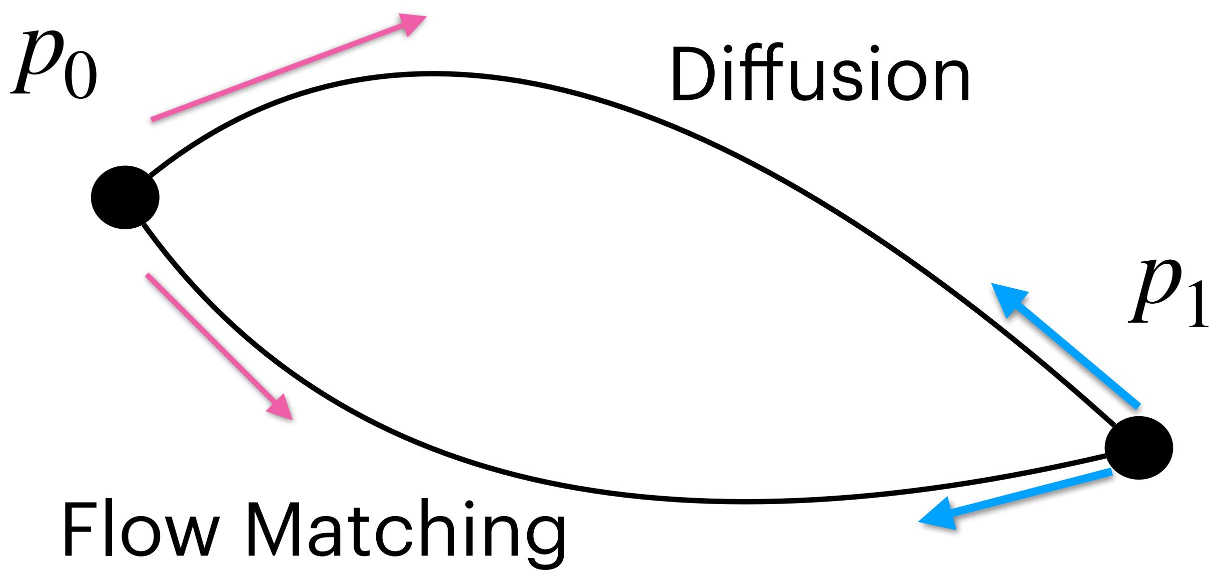
1. Choose $f(\cdot)$ and $g_t(\cdot)$

Learn the reverse process

1. $b(\cdot) := f - g^2 \nabla_x \log p(\cdot)$

Fokker Planck Equation

$$\frac{\partial p_t}{\partial t} = -\frac{\partial}{\partial x}[\boxed{f(x,t)}p(x,t)] + \frac{\partial^2}{\partial x^2}[\frac{1}{2}\boxed{g^2}p(x,t)]$$



Flow Matching:

Predefine data pairing $(x_0, x_1) \sim p_0 \otimes p_1$

1. Construct linear interpolation vector field $f(\cdot, t \,|\, x_0, x_1) := x_1 - x_0$, $g_t \equiv 0$

Learn the reverse process

1. $b(\cdot) := -f(\cdot)$

Then what is the take away here?

Design Space of Dynamical Generative Modeling

1. Take aways

- The path measure between p_0 and p_1 is not unique.
- Regressing the drift term.

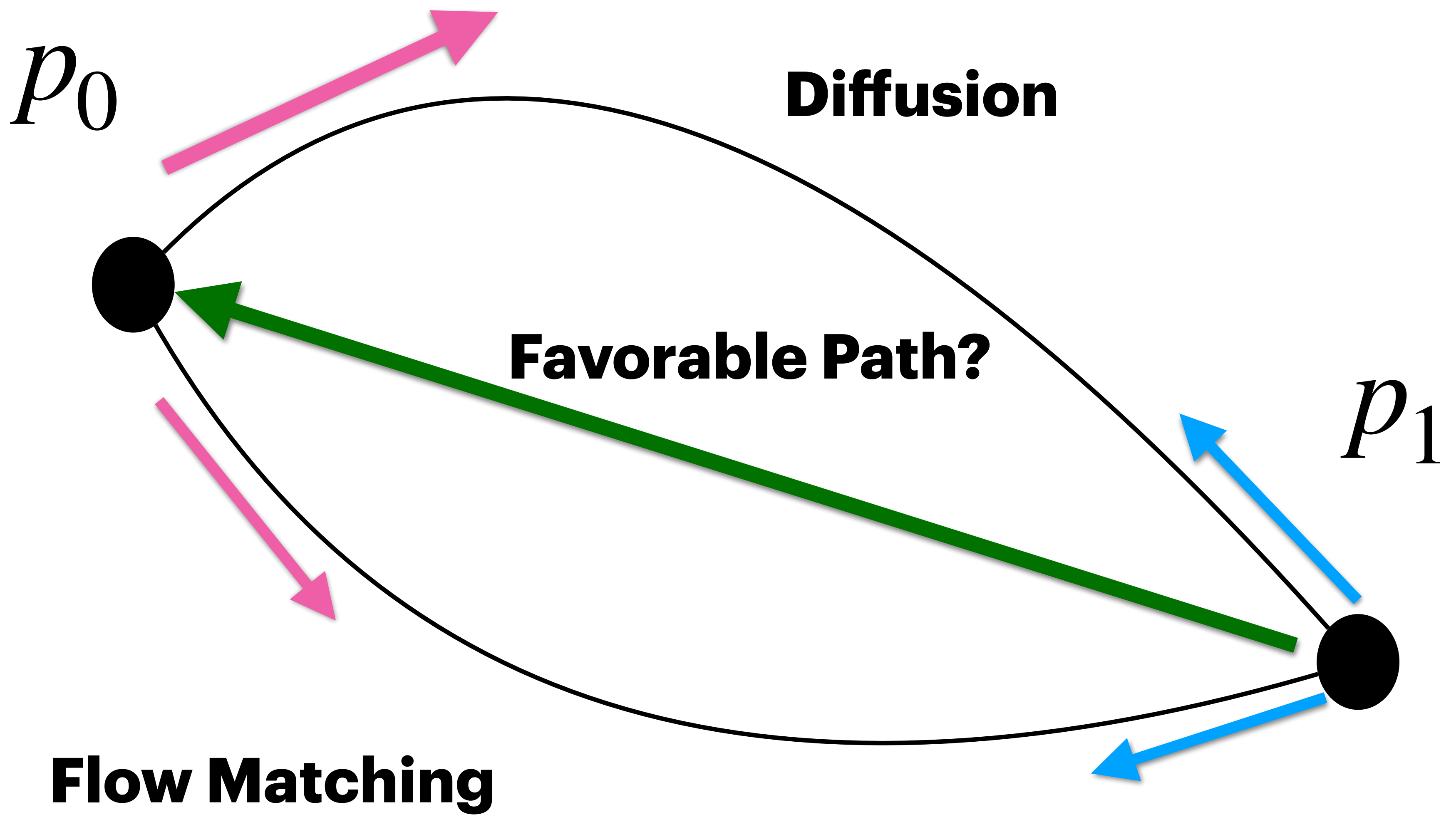
2. Design Space and motivations

Q: What is the favorable Path?

A: Task Specific:

1. **Generative Modeling:** smooth and straight path For fast sampling.

2. **Trajectory Inference:** 'Natural' Interpolation between distributions $p_0, p_1 ..., p_N$.



Acceleration Generative Model

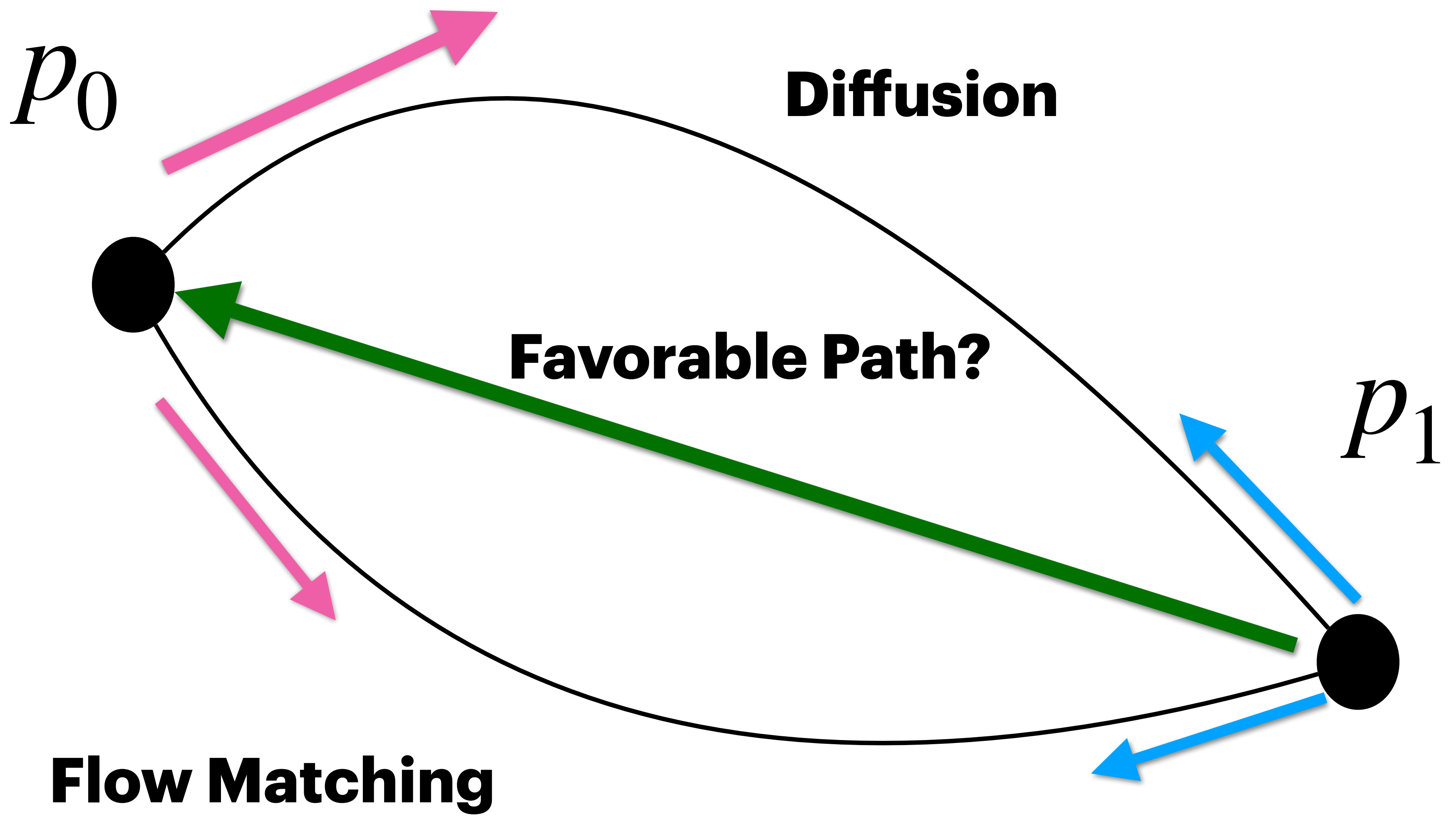
Design Space of Dynamical Generative Modeling

Task:

1. **Generative Modeling:** smooth and straight path For fast sampling.

2. Design Space and motivations

- Can we incorporate more information of dynamics (i.e Velocity)?
- Construct unique path measure which is favorable for fast sampling.



Phase Space Dynamics

$$dx_t = b(x_t, t)dt + g_t dw_t$$

$x_0 \sim p_0$

$x_1 \sim p_1$

- We would like to track the <u>auxiliary velocity variable</u> in the dynamics in order to further utilize them.

Phase Space Dynamics

$$dx_t = v_t dt$$

$$dv_t = a(x_t, v_t, t)dt + g_t dw_t$$

$$\cancel{dx_t = b(x_t, t)dt + g_t dw_t}$$

$x_0 \sim p_0$

$x_1 \sim p_1$

- We use Phase space dynamics (Newtonian Dynamics [1]) in which we have <u>auxiliary velocity variable</u> in the dynamics, similar to Critically Damped Langevin Dynamics ([2],CLD).

[1] https://en.wikipedia.org/wiki/Newtonian_dynamics
[2] Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion." *arXiv preprint arXiv:2112.07068* (2021).

Phase Space Dynamics

$$dx_t = v_t dt$$
$$dv_t = a(x_t, v_t, t)dt + g_t dw_t$$

$$\rightleftarrows$$

$$dm_t = F_t dt + G_t dW_t \quad m = [x_t, v_t]^T$$

$x_0 \sim p_0$

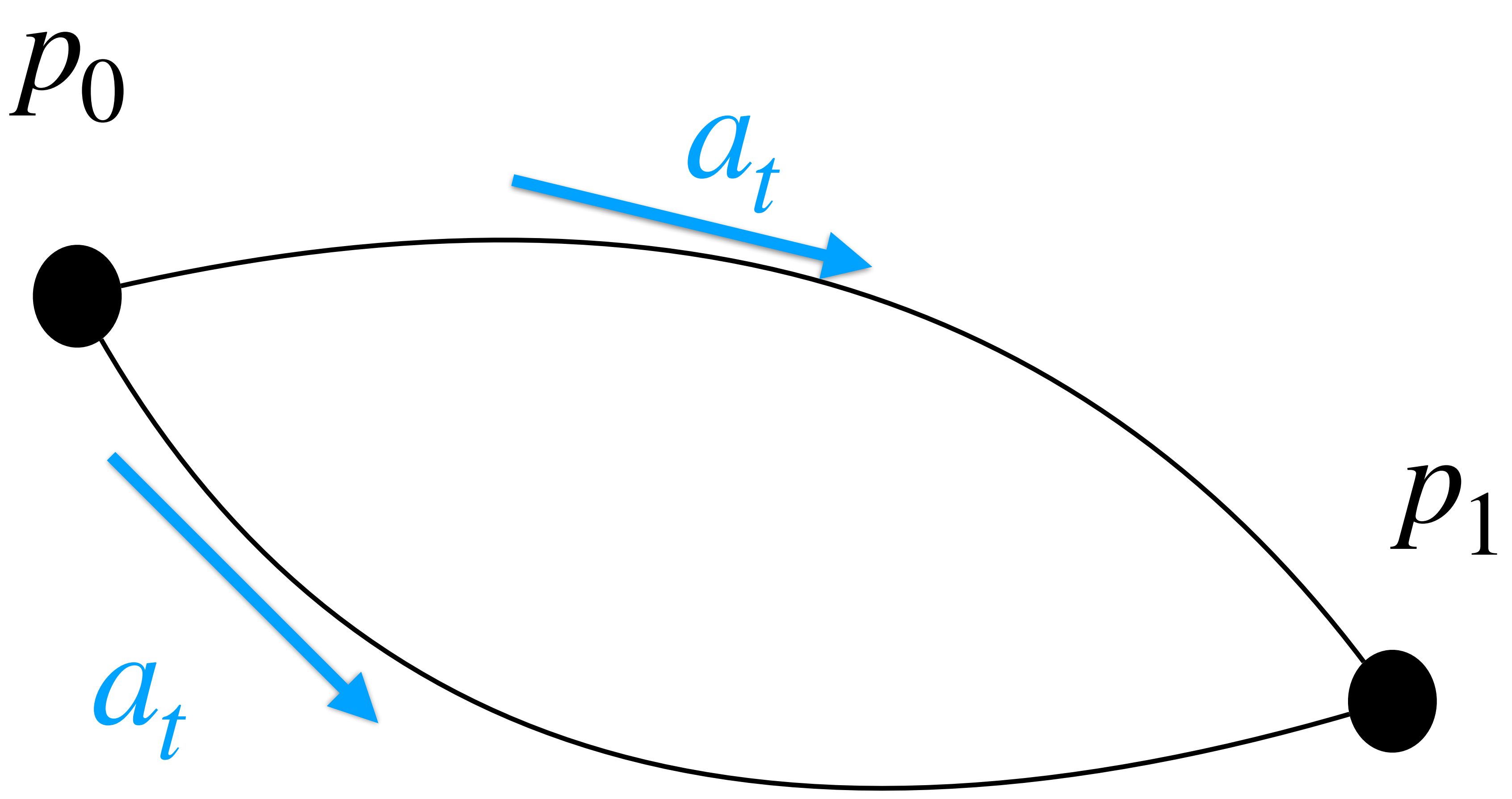
$x_1 \sim p_1$

- We use Phase space dynamics (Newtonian Dynamics [1]) in which we have auxiliary velocity variable in the dynamics, similar to Critically Damped Langevin Dynamics ([2],CLD).

[1] Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion." *arXiv preprint arXiv:2112.07068* (2021).

Optimal Control Phase Space Dynamics

$$dx_t = v_t dt$$
$$dv_t = a(x_t, v_t, t)dt + g_t dw_t$$



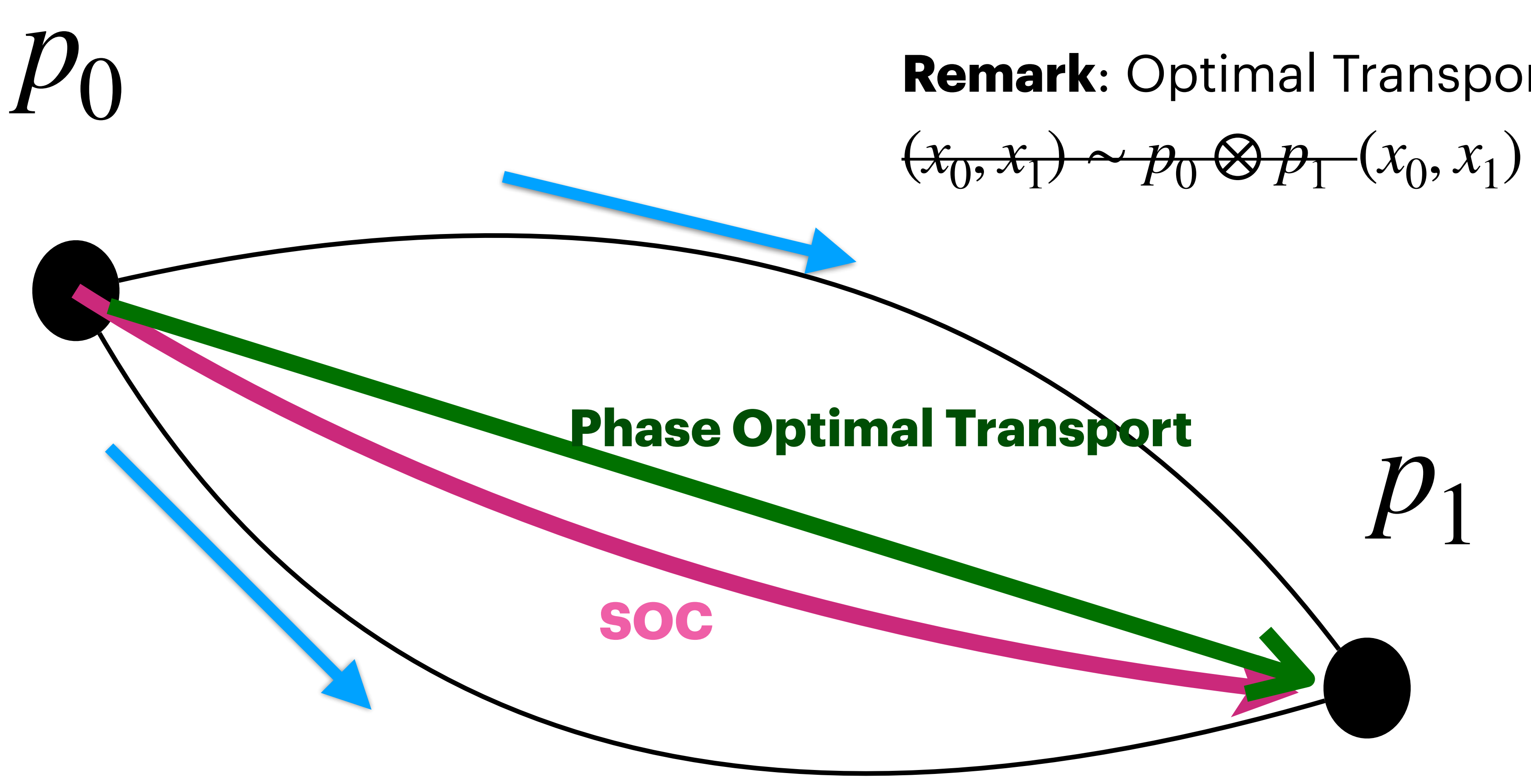
- **Q:**However, <u>The trajectory is still not unique since a_t is not unique .</u> Which path measure are we looking for?
- **A:**We hope the path is straight and smooth in the position and velocity space! The **sub-optimal** drift term is the **<u>solution of a Stochastic Optimal Control (SOC) Problem</u>**!

$$a^* = argmin_a \mathbb{E}\left[\int_0^1 ||a_t(x_t, v_t, t)||_2^2 dt\right]$$

s.t
$$dx_t = v_t dt$$
$$dv_t = a(x_t, v_t, t)dt + g_t dw_t$$
$$(x_0, x_1) \sim p_0 \otimes p_1$$

Optimal Control Phase Space Dynamics

$$dx_t = v_t dt$$
$$dv_t = a(x_t, v_t, t)dt + g_t dw_t$$



Remark: Optimal Transport needs correct pairing
$\cancel{(x_0, x_1) \sim p_0 \otimes p_1}\ (x_0, x_1) \sim \Pi^{OT}(p_0, p_1)$

- **Q:** However, The trajectory is still not unique since a_t is not unique . Which path measure are we looking for?
- **A:** We hope the path is straight and smooth in the position and velocity space! The **sub-optimal** drift term is the **solution of a Stochastic Optimal Control (SOC) Problem**!

$$a^* = g_t^2 P_{11}\left[\frac{x_1 - x_t}{1 - t} - v_t\right]$$

$$P_{11} := \frac{-4}{g_t^2(t-1)}$$

Optimal Control Phase Space Dynamics

$x_0 \sim p_0$

$x_1 \sim p_1$

$$\begin{bmatrix} \mathrm{d}\mathbf{x}_t \\ \mathrm{d}\mathbf{v}_t \end{bmatrix} = \begin{bmatrix} \mathbf{v}_t \\ \mathbf{F}_t \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{bmatrix} \mathrm{d}\mathbf{w}_t \quad \text{s.t} \quad \mathbf{m}_0 := \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{v}_0 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0),$$

$$\text{Bridge Matching SDE} : \mathbf{F}_t := \mathbf{F}_t^b(\mathbf{m}_t, t) \equiv \mathbf{a}_t^*(\mathbf{m}_t, t),$$

$$\text{Probablistic ODE} : \mathbf{F}_t := \mathbf{F}_t^p(\mathbf{m}_t, t) \equiv \mathbf{a}_t^*(\mathbf{m}_t, t) - \frac{1}{2} g_t^2 \nabla_{\mathbf{v}} \log p(\mathbf{m}, t)$$

Quality

Speed

- **Training:** One can regress F_t using neural network to construct the SDE or ODE to make the bridge between two distributions.

$$L = \mathbb{E}_t \mathbb{E}_{x_0} \mathbb{E}_{x_t} || F_t^\theta(x_t, v_t, t) - F_t ||_2^2$$

Wait, why it is better?

Design Space of Dynamical Generative Modeling

Task:

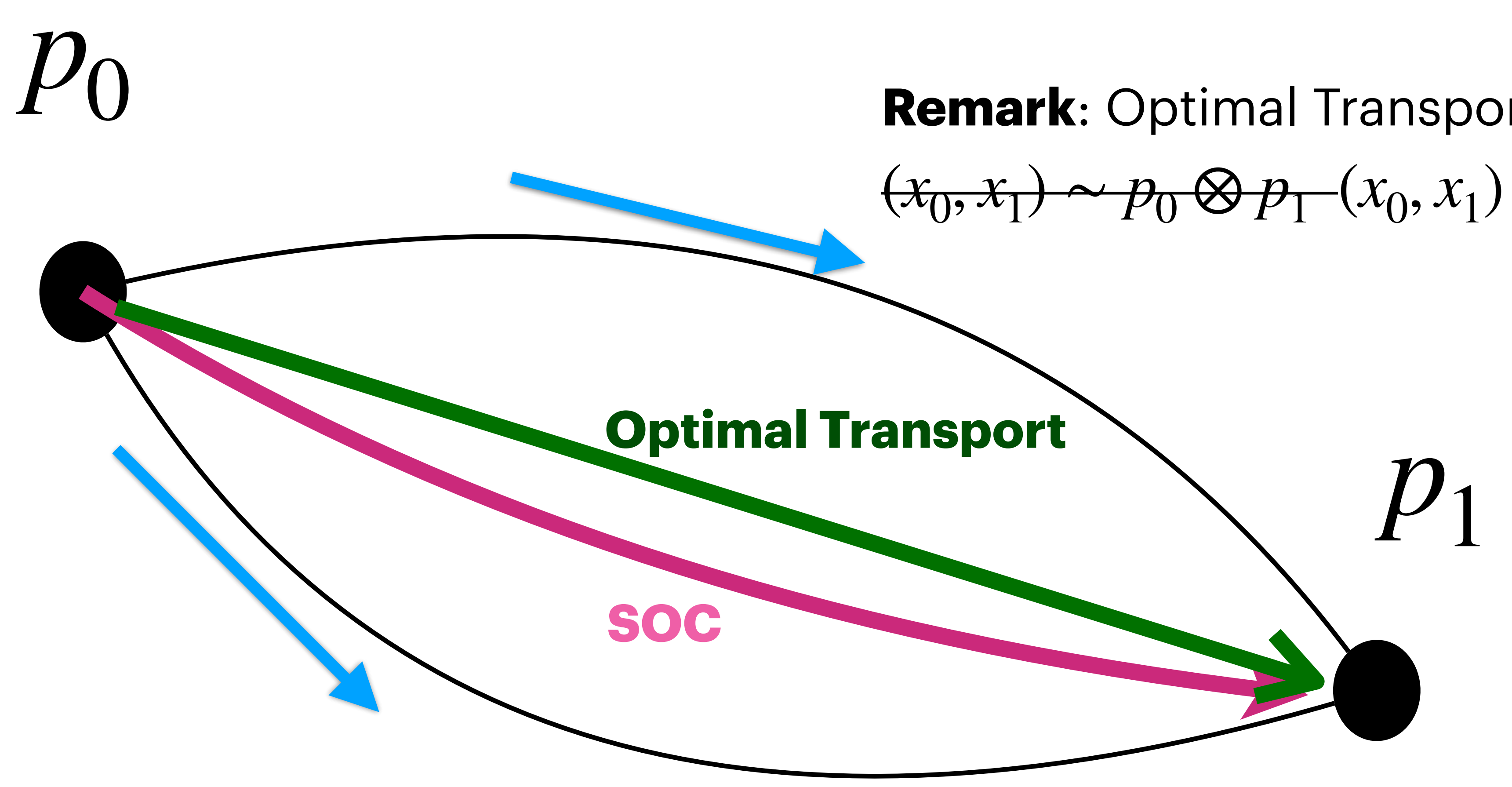
1. **Generative Modeling:** smooth and straight path For fast sampling.

2. Design Space and motivations

- Can we incorporate <u>more information</u> of dynamics (i.e Velocity)?
- Construct <u>unique path measure</u> which is <u>favorable for fast sampling.</u>

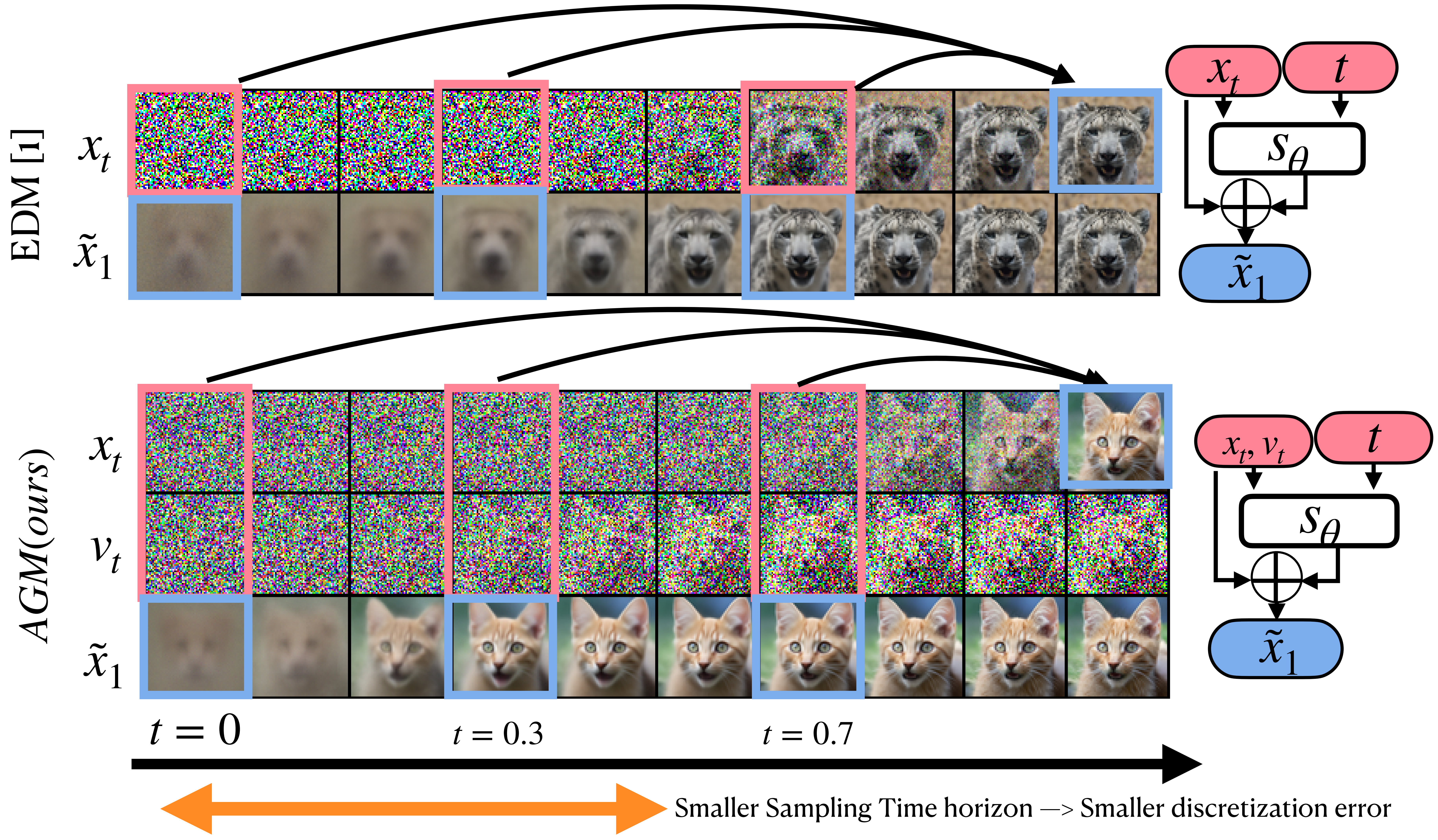
Remark: Optimal Transport needs correct pairing

$\cancel{(x_0, x_1) \sim p_0 \otimes p_1}\ (x_0, x_1) \sim \Pi^{OT}(p_0, p_1)$



Sampling Hop

With additional information of the dynamics



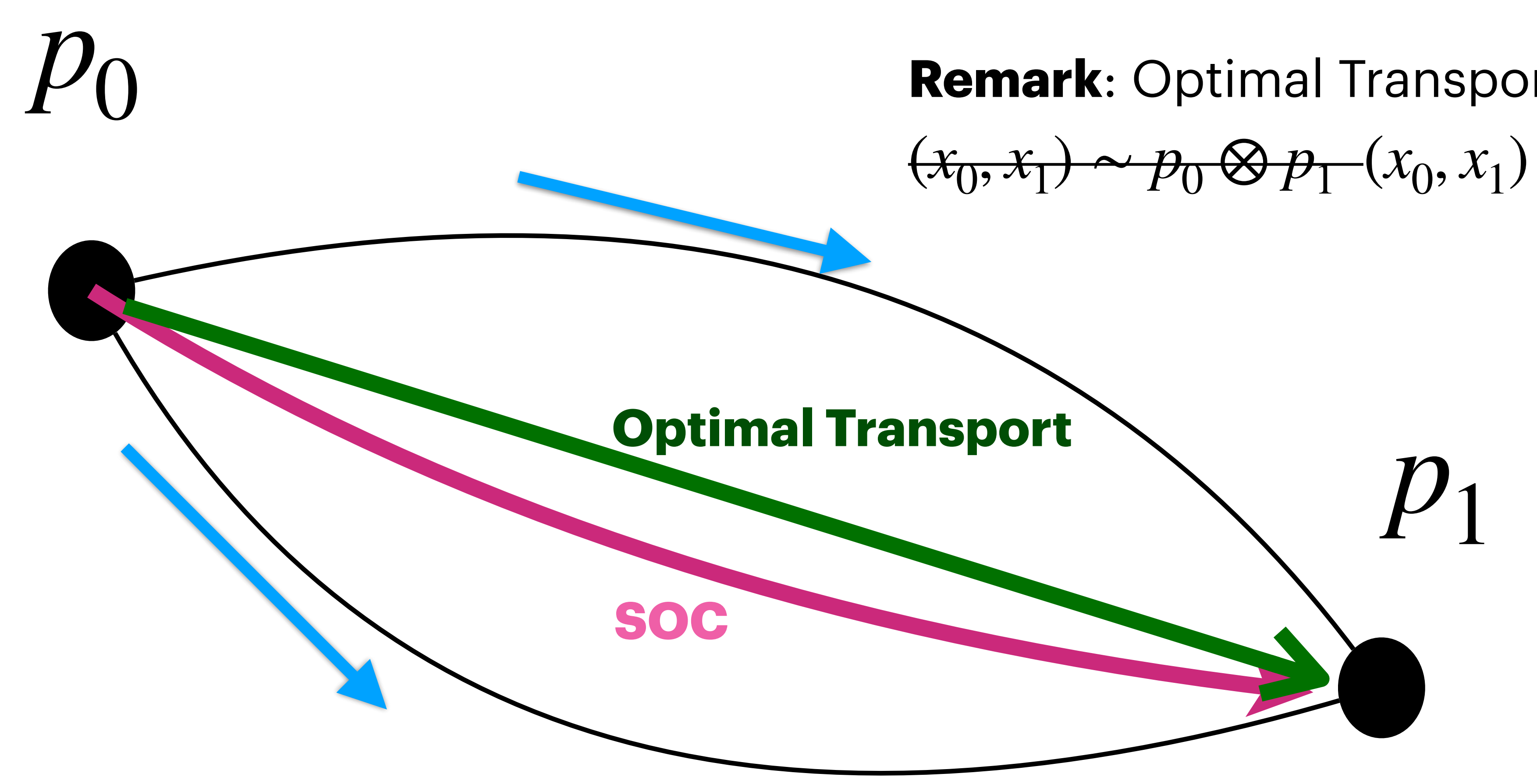
[1]Karras, Tero, et al. "Elucidating the design space of diffusion-based generative models." *Advances in Neural Information Processing Systems* 35 (2022): 26565-26577.

Design Space of Dynamical Generative Modeling

1. Task:

1. **Generative Modeling:** smooth and straight path For fast sampling.

Remark: Optimal Transport needs correct pairing
$\cancel{(x_0, x_1) \sim p_0 \otimes p_1}$ $(x_0, x_1) \sim \Pi^{OT}(p_0, p_1)$

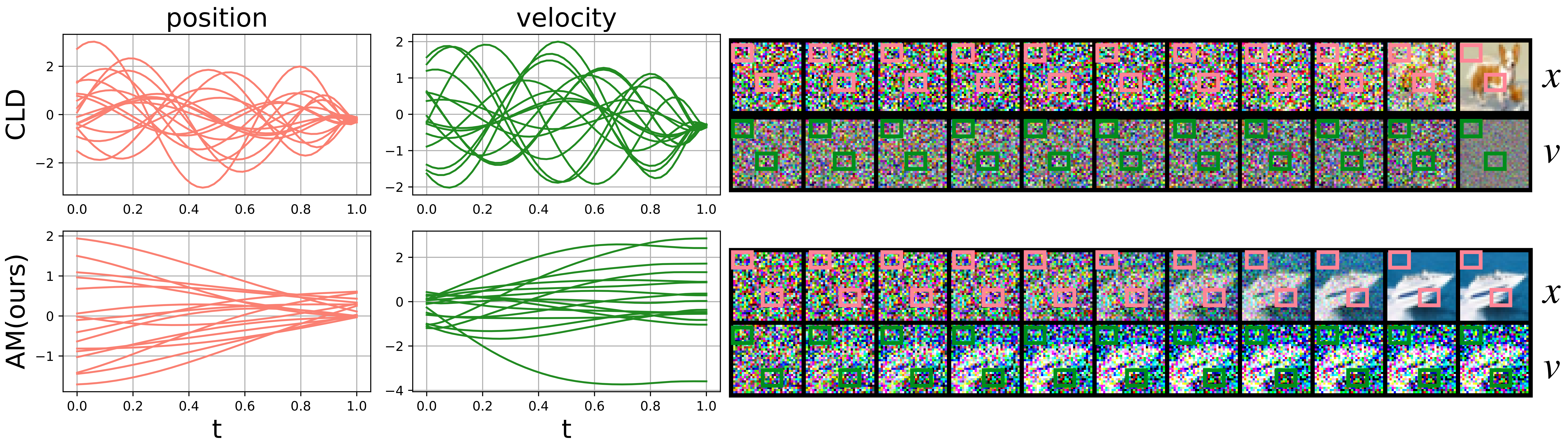


2. Design Space and motivations

- Let model incorporate <u>more information</u> of dynamics (i.e Velocity).
- Construct <u>unique path measure</u> which is <u>favorable for sampling</u>

Stochastic Optimal Control
Induces Smoother and Straighter Trajectories

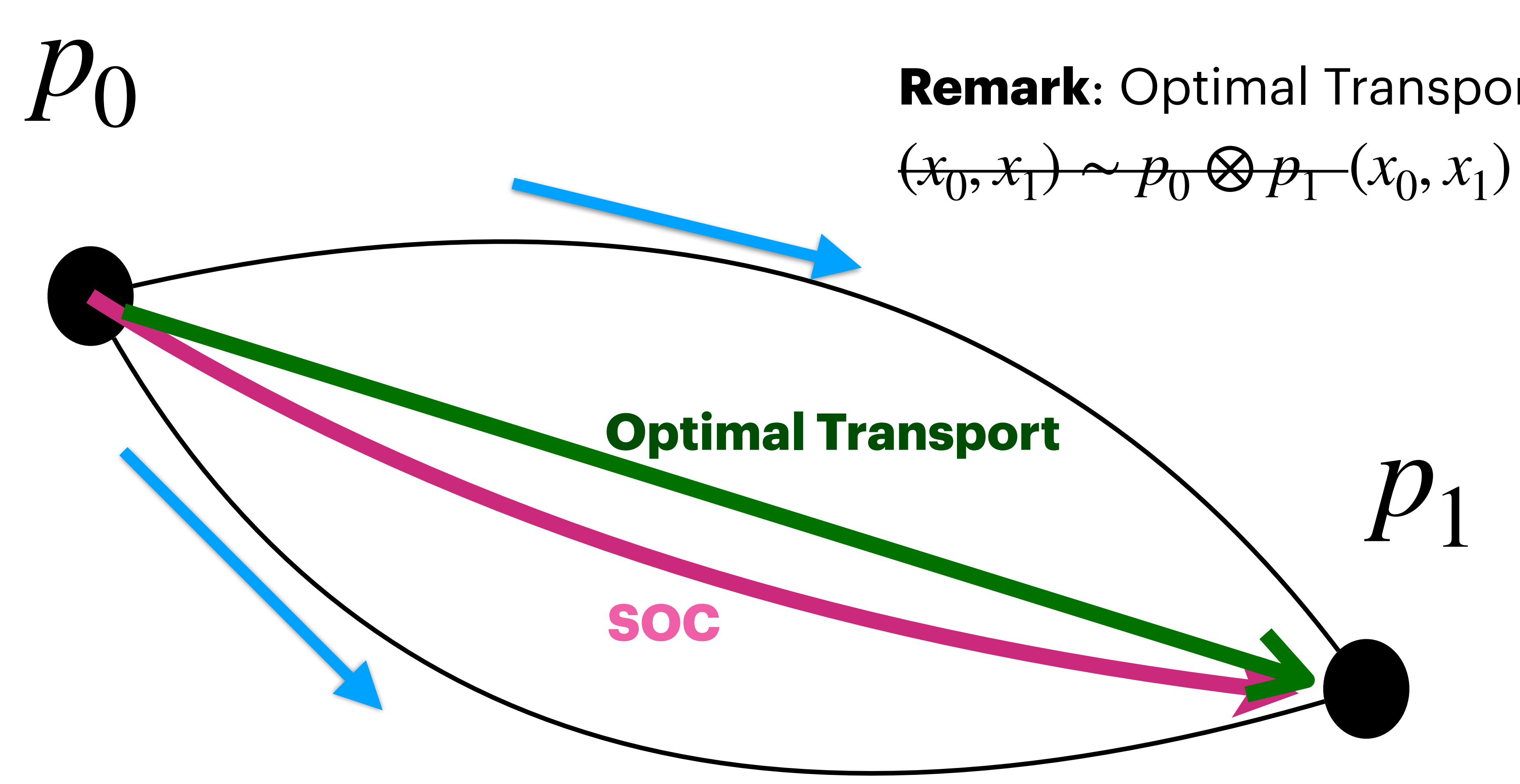
Trajectories of random sampled pixels compared with Critically Damped Langevin Dynamics.



Design Space of Dynamical Generative Modeling

Task:

1. **Generative Modeling:** smooth and straight path For fast sampling.



Remark: Optimal Transport needs correct pairing

$\cancel{(x_0, x_1) \sim p_0 \otimes p_1}$ $(x_0, x_1) \sim \Pi^{OT}(p_0, p_1)$

2. Design Space and motivations

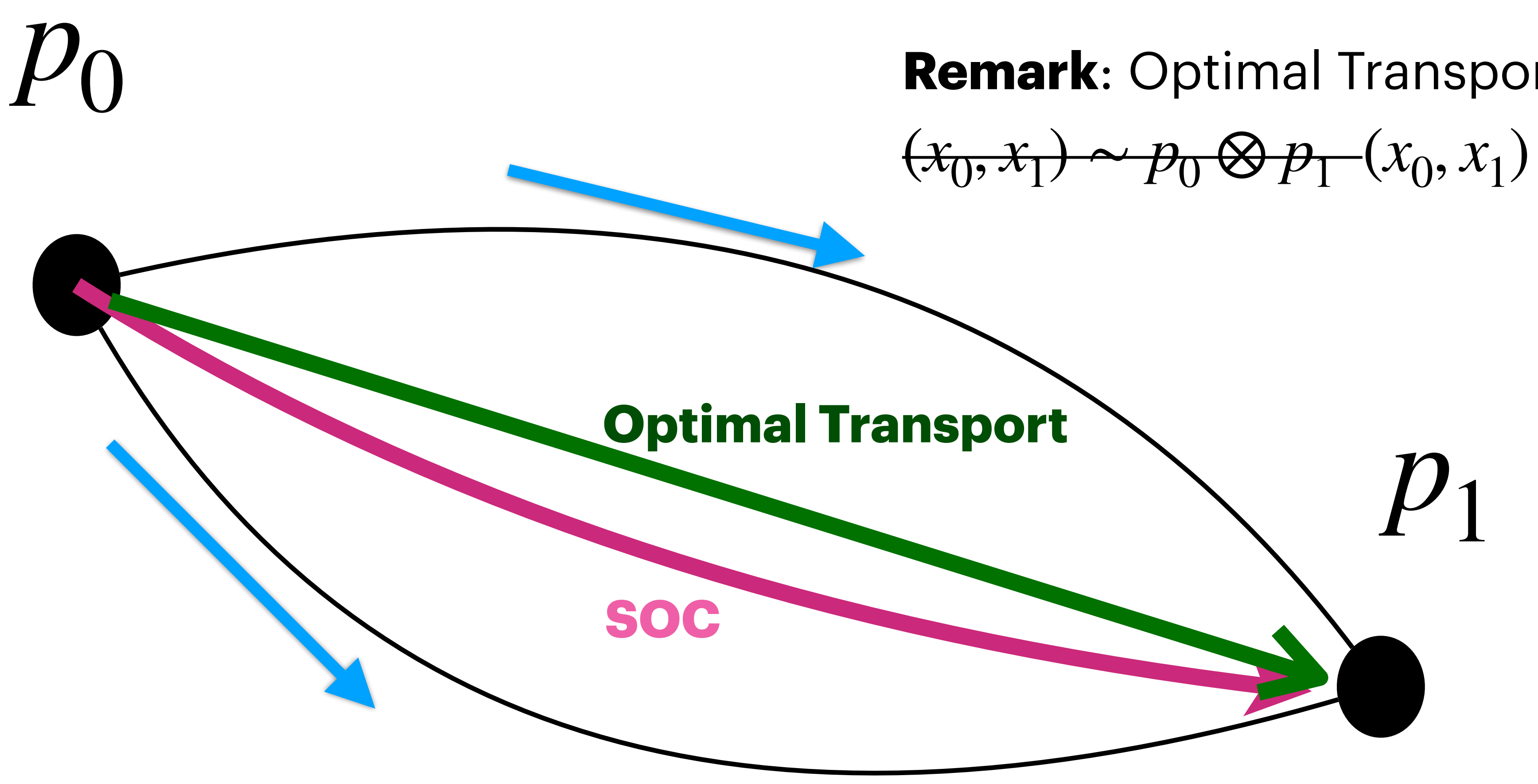
- Let model incorporate more information of dynamics (i.e Velocity).
- Construct unique path measure which is favorable for sampling

That's it?

Design Space of Dynamical Generative Modeling

Task:

1. **Generative Modeling:** smooth and straight path For fast sampling.



Remark: Optimal Transport needs correct pairing
$\cancel{(x_0, x_1) \sim p_0 \otimes p_1}\ (x_0, x_1) \sim \Pi^{OT}(p_0, p_1)$

2. Design Space and motivations

- Let model incorporate more information of dynamics (i.e Velocity).
- Construct unique path measure which is favorable for sampling

} **That's it? No!**

Delay Error in Momentum System

1. We find that there will be Delay Error When conducting sampling (propagating momentum dynamics.)

Euler Discretization:

When $t = t$

$$x_{t+\delta_t} = x_t + v_t \cdot \delta_t$$

$$v_{t+\delta_t} = v_t + F_t^\theta(x_t, v_t, t) \cdot \delta_t$$

When $t = t + \delta_t$

$$x_{t+2\delta_t} = x_{t+\delta_t} + v_{t+\delta_t} \cdot \delta_t$$

$$v_{t+2\delta_t} = v_{t+\delta_t} + F_{t+\delta_t}^\theta(x_{t+\delta_t}, v_{t+\delta_t}, t + \delta_t) \cdot \delta_{t+\delta_t}$$

Learnt Force term F_t^θ cannot influence the position until next timestep. It is obvious that such error will be amplified when interval δ_t is large.

Delay Error in Momentum System

Euler Discretization:

When $t = t$

$$x_{t+\delta_t} = x_t + v_t \cdot \delta_t$$

$$v_{t+\delta_t} = v_t + F_t^\theta(x_t, v_t, t) \cdot \delta_t$$

When $t = t + \delta_t$

$$x_{t+2\delta_t} = x_{t+\delta_t} + v_{t+\delta_t} \cdot \delta_t$$

$$v_{t+2\delta_t} = v_{t+\delta_t} + F_{t+\delta_t}^\theta(x_{t+\delta_t}, v_{t+\delta_t}, t + \delta_t) \cdot \delta_{t+\delta_t}$$

Such system is <u>well-known under-actuated system</u> in the control/robotics domain, in which the control/Force cannot be injected into full channel.

Delay Error in Momentum System

Exponential Integrator:

Given System: $dx_t = (\boxed{A(t)}x_t + \boxed{B(t)}F_t)dt$

The solution is: $x(t) = \Phi(t,0)x_0 + \int_0^t \phi(t,\tau)B(t)F_\tau d\tau$

Where: $\Phi(\cdot\,,\cdot)$ is the transition kernel of <u>uncontrolled system.</u>

Maybe Unknown Fact:

DDIM[1] is <u>*EXACTLY*</u> same as Exponential Integrator [2] in continuous limit.

[1] Song, Jiaming, Chenlin Meng, and Stefano Ermon. "Denoising diffusion implicit models." *arXiv preprint arXiv:2010.02502* (2020).
[2]Qinsheng Zhang et al. fast sampling of Diffusion Models with Exponential Integrator

Delay Error in Momentum System

Exponential Integrator:

$$\begin{bmatrix}\mathbf{x}_{t_{i+1}}\\ \mathbf{v}_{t_{i+1}}\end{bmatrix} = \Phi(t_{i+1}, t_i)\begin{bmatrix}\mathbf{x}_t\\ \mathbf{v}_t\end{bmatrix} + \sum_{j=0}^{w}\begin{bmatrix}\int_{t_i}^{t_{i+1}}(t_{i+1}-\tau)\,\mathbf{z}_\tau\cdot\mathbf{M}_{i,j}(\tau)\mathrm{d}\tau\cdot\boxed{\mathbf{s}_t^\theta(\mathbf{m}_{t_{i-j}}, t_{i-j}))}\\ \int_{t_i}^{t_{i+1}}\mathbf{z}_\tau\cdot\mathbf{M}_{i,j}(\tau)\mathrm{d}\tau\cdot\boxed{\mathbf{s}_t^\theta(\mathbf{m}_{t_{i-j}}, t_{i-j})}\end{bmatrix}$$

$$\text{Where } \mathbf{M}_{i,j}(\tau) = \prod_{k\neq j}\left(\frac{\tau - t_{i-k}}{t_{i-j}-t_{i-k}}\right), \quad \text{and} \quad \Phi(t,s) = \begin{bmatrix}1 & t-s\\ 0 & 1\end{bmatrix}.$$

Where we reparamterized: $F_t^\theta := s_t^\theta \cdot z_t$, and z_t is the normalizer which normalizes the output of neural network s_t^θ to standard variance.

Delay Error in Momentum System

Exponential Integrator:

Mitigate Delay issue

$$\begin{bmatrix}\mathbf{x}_{t_{i+1}}\\ \mathbf{v}_{t_{i+1}}\end{bmatrix} = \Phi(t_{i+1}, t_i)\begin{bmatrix}\mathbf{x}_t\\ \mathbf{v}_t\end{bmatrix} + \sum_{j=0}^{w}\begin{bmatrix}\int_{t_i}^{t_{i+1}}(t_{i+1}-\tau)\,\mathbf{z}_\tau\cdot\mathbf{M}_{i,j}(\tau)\mathrm{d}\tau\cdot\boxed{\mathbf{s}_t^\theta(\mathbf{m}_{t_{i-j}}, t_{i-j})}\\ \int_{t_i}^{t_{i+1}}\mathbf{z}_\tau\cdot\mathbf{M}_{i,j}(\tau)\mathrm{d}\tau\cdot\boxed{\mathbf{s}_t^\theta(\mathbf{m}_{t_{i-j}}, t_{i-j})}\end{bmatrix}$$

$$\text{Where } \mathbf{M}_{i,j}(\tau) = \prod_{k\neq j}\left(\frac{\tau - t_{i-k}}{t_{i-j}-t_{i-k}}\right), \quad \text{and} \quad \Phi(t,s) = \begin{bmatrix}1 & t-s\\ 0 & 1\end{bmatrix}.$$

Where we reparamterized: $F_t^\theta := s_t^\theta \cdot z_t$, and z_t is the normalizer which normalizes the output of neural network s_t^θ to standard variance.

Acceleration Generative Model

Components

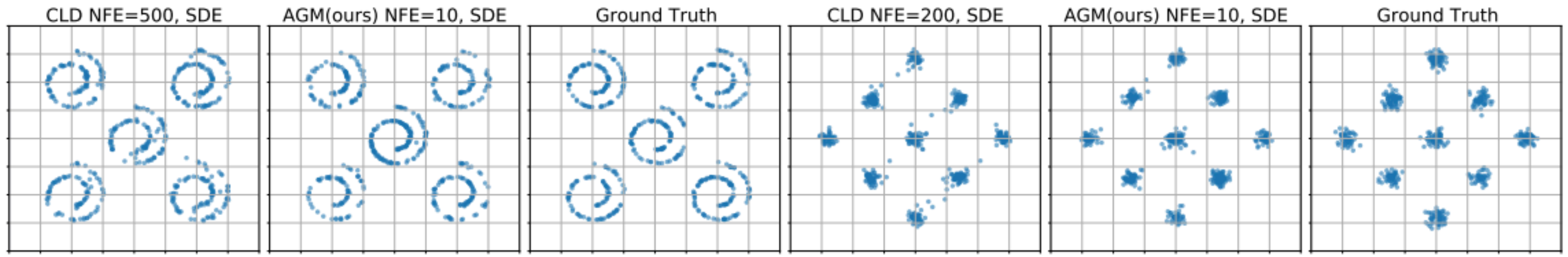
1. Stochastic Phase Space Dynamics
2. Stochastic Optimal Control
3. Exponential Integrator

Algorithm Design

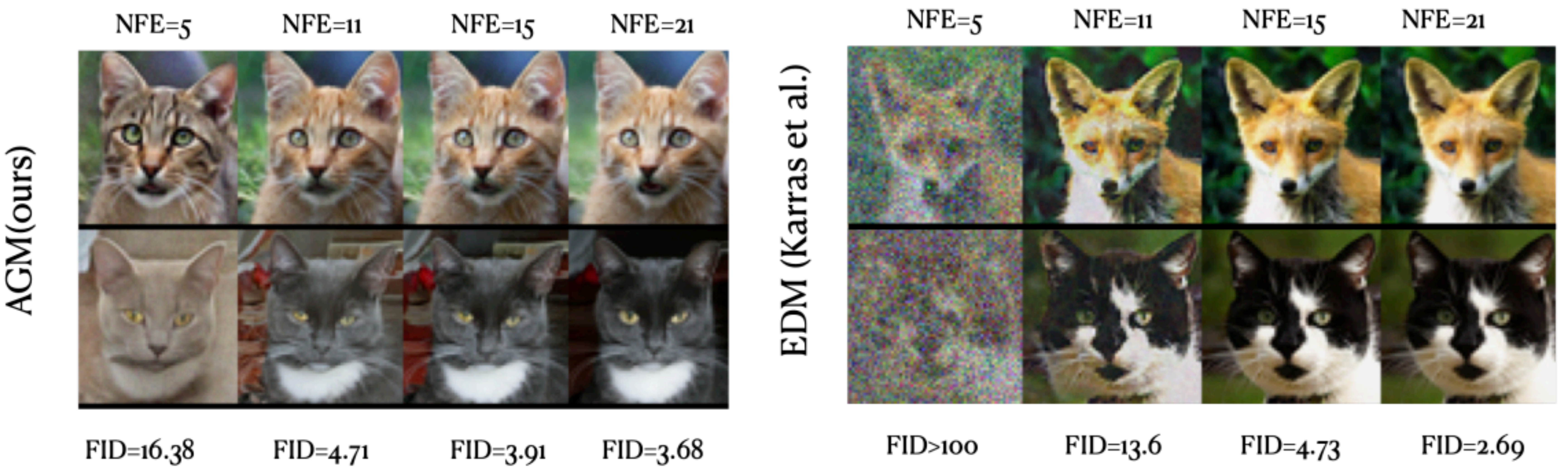
1. Sampling-Hop
2. Straighten Trajectories
3. Resolve delay issue

Experimental Results

Compare with CLD[1]:



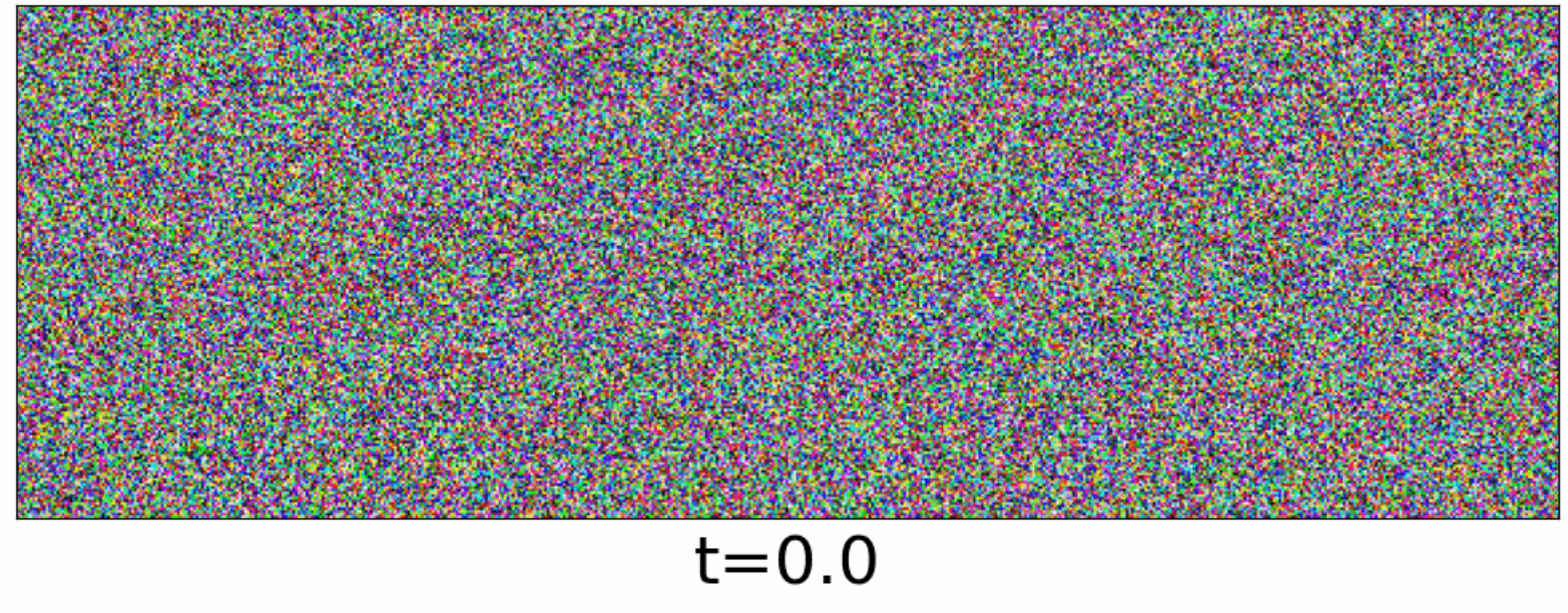
Compare with EDM [2]:



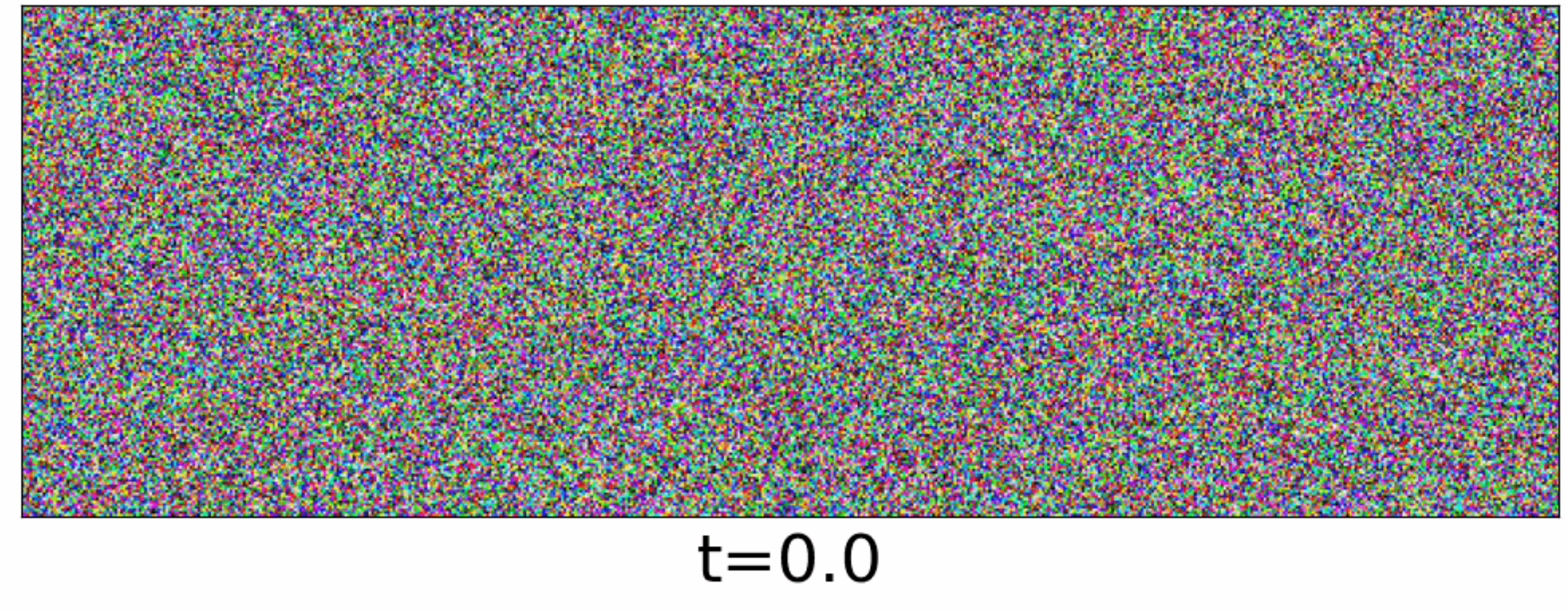
[1]Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion." *arXiv preprint arXiv:2112.07068* (2021).
[2]Karras, Tero, et al. "Elucidating the design space of diffusion-based generative models." *Advances in Neural Information Processing Systems* 35 (2022): 26565-26577.

Experimental Results

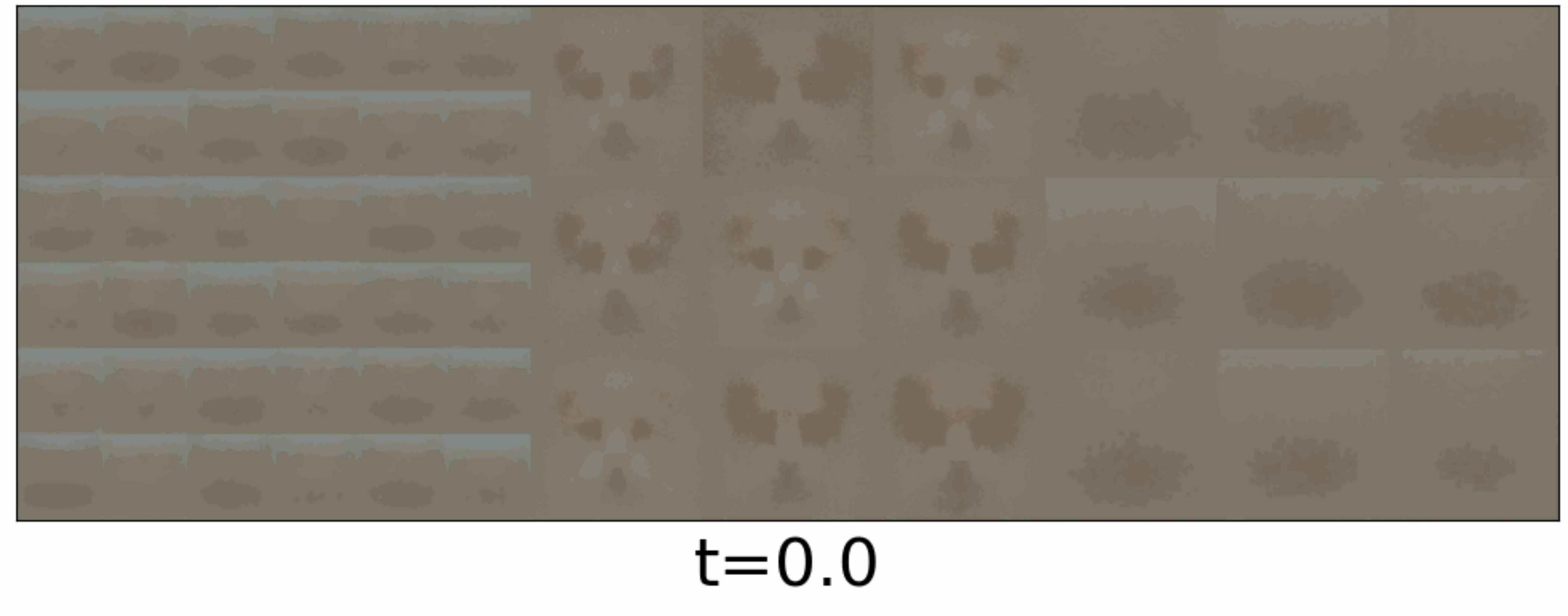
Position



Velocity



Estimated data



Unconditional ImageNet-64x64:

NFE=40, FID 10.97

NFE=20, FID 12.55

3x faster <— minibatch OT FM

Better performance

No OT pairing

Unconditional Cifar10:

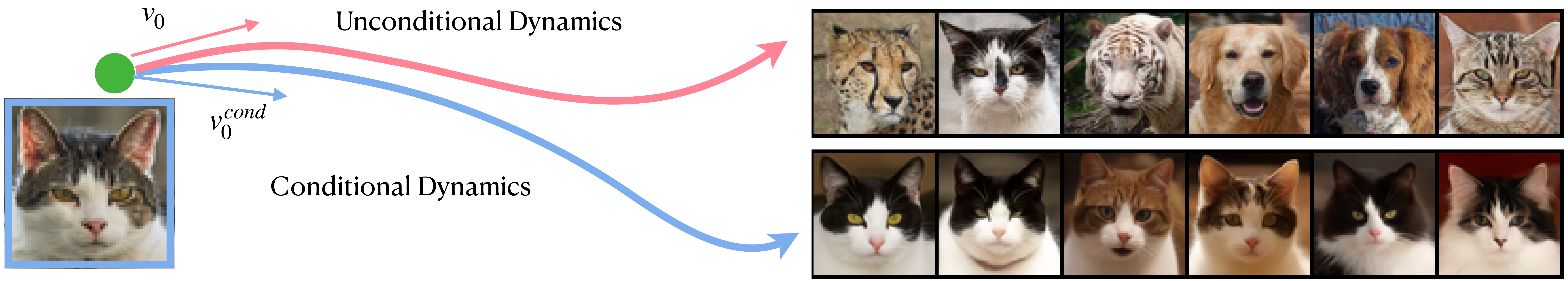
NFE=5, FID=11.93

NFE=10, FID=4.6

NFE=20,FID=2.6

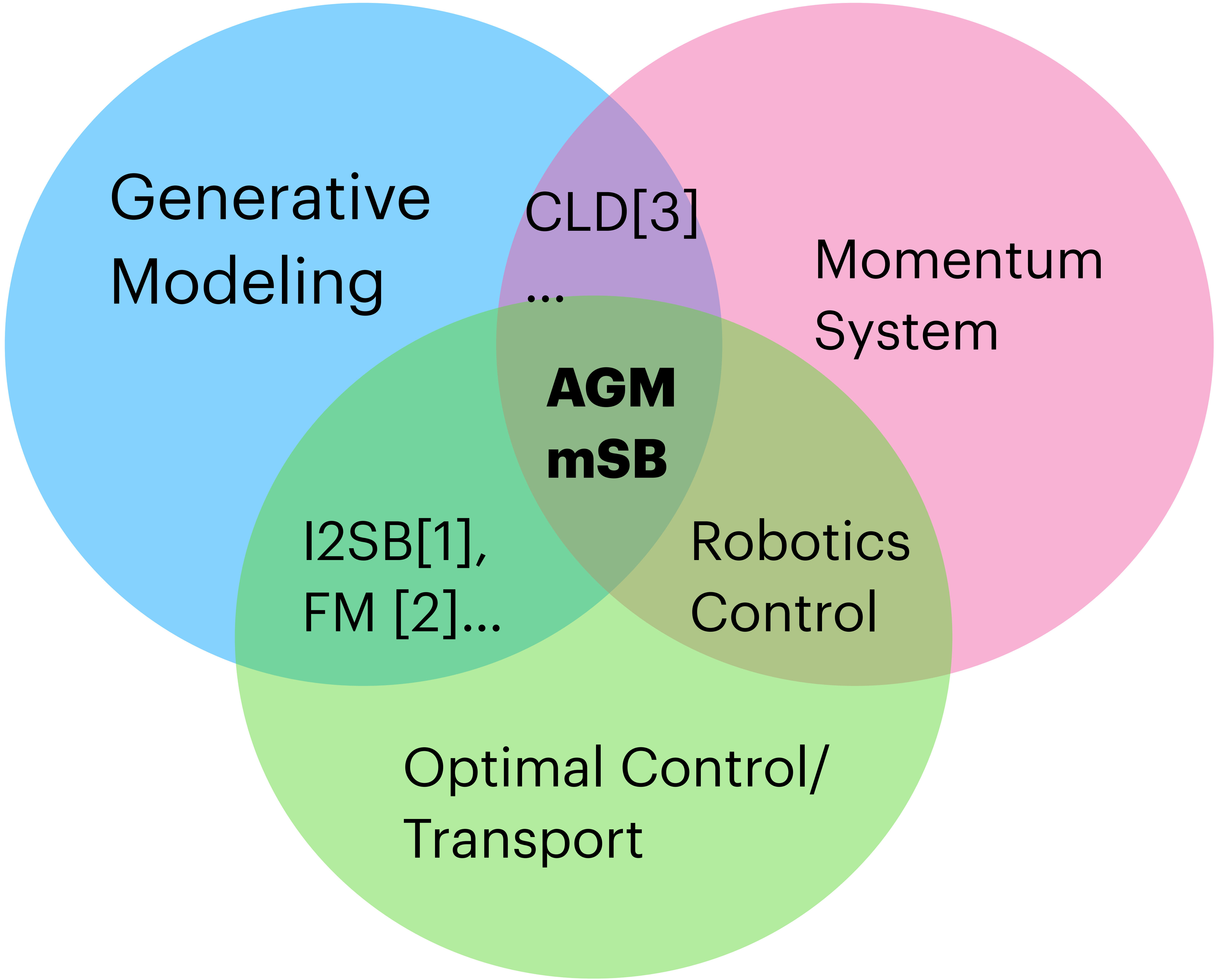
Stroke Based Conditional Generation

By leveraging velocity Space



1. Velocity has explicitly physical meaning as momentum.
2. One can guide the generation path by only changing initial auxiliary velocity.
3. The model does not need fine tuning or further training.

Summary



AGM:

1. **Momentum System:** extra information for sampling-hop.
2. **Optimal control:**
    1. straight trajectory.
    2. Exponential Integrator

[1] Liu, Guan-Horng, et al. "I $^ 2$ SB: Image-to-Image Schr\" odinger Bridge." *arXiv preprint arXiv:2302.05872* (2023).
[2] Lipman, Yaron, et al. "Flow matching for generative modeling." *arXiv preprint arXiv:2210.02747* (2022).
[3] Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion."

Momentum Schrödinger Bridge

Design Space of Dynamical Generative Modeling

1. Task:

1. **Trajectory Inference:** ‘Natural’ Interpolation between distributions.

2. Why and what is ‘natural’?

1. Some system follow optimality principle.

2. Example: minimum control (least effort) in the evolution of the Single Cell.

\

Design Space of Dynamical Generative Modeling

1. Task:

1. **Trajectory Inference:** 'Natural' Interpolation between distributions.

	Models	Optimality	$p_0(\cdot)$	$p_1(\cdot)$
Generative Modeling	SGM [1]	✗	$p_{\mathcal{A}}(x)$	$\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
	CLD [2]	✗	$p_{\mathcal{A}}(x) \otimes \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$	$\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \otimes \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
Trajectory Inference	SB [3]	$\downarrow W_2 \rightarrow$ kinks	$p_{\mathcal{A}}(x)$	$p_{\mathcal{B}}(x)$
	DMSB (ours)	$\downarrow W_2 \rightarrow$ smooth	$p_{\mathcal{A}}(x) q_\theta(v\|x)$	$p_{\mathcal{B}}(x) q_\phi(v\|x)$

2. Why and what is 'natural'?

1. Some system follow optimality principle.

2. Example: minimum control (least effort) in the evolution of the Single Cell.

[1]Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).
[2]Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion."
[3]Chen, Tianrong, Guan-Horng Liu, and Evangelos A. Theodorou. SB-FBSDE

Design Space of Dynamical Generative Modeling

	Models	Optimality	$p_0(\cdot)$	$p_1(\cdot)$
Generative Modeling	SGM [1]	✗	$p_{\mathcal{A}}(x)$	$\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
	CLD [2]	✗	$p_{\mathcal{A}}(x) \otimes \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$	$\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \otimes \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
Trajectory Inference	SB [3]	$\downarrow W_2 \rightarrow$ kinks	$p_{\mathcal{A}}(x)$	$p_{\mathcal{B}}(x)$
	DMSB (ours)	$\downarrow W_2 \rightarrow$ smooth	$p_{\mathcal{A}}(x) q_\theta(v\|x)$	$p_{\mathcal{B}}(x) q_\phi(v\|x)$

1. Task:

1. **Trajectory Inference:** 'Natural' Interpolation between distributions.

2. Why and what is 'natural'?

1. Some system follow optimality principle.

2. Example: least effort in the evolution of the Single Cell sequence.

3. Design Space and motivations ("Natural")

- Newtonian Dynamics (follow Newton's law of motion).
- Principle of least effort. (biology).
- Evolution is stochastic.

[1]Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).
[2]Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion."
[3]Chen, Tianrong, Guan-Horng Liu, and Evangelos A. Theodorou. SB-FBSDE

Design Space of Dynamical Generative Modeling

1. Task:

1. **Trajectory Inference:** ‘Natural’ Interpolation between distributions.

2. Why and what is ‘natural’?

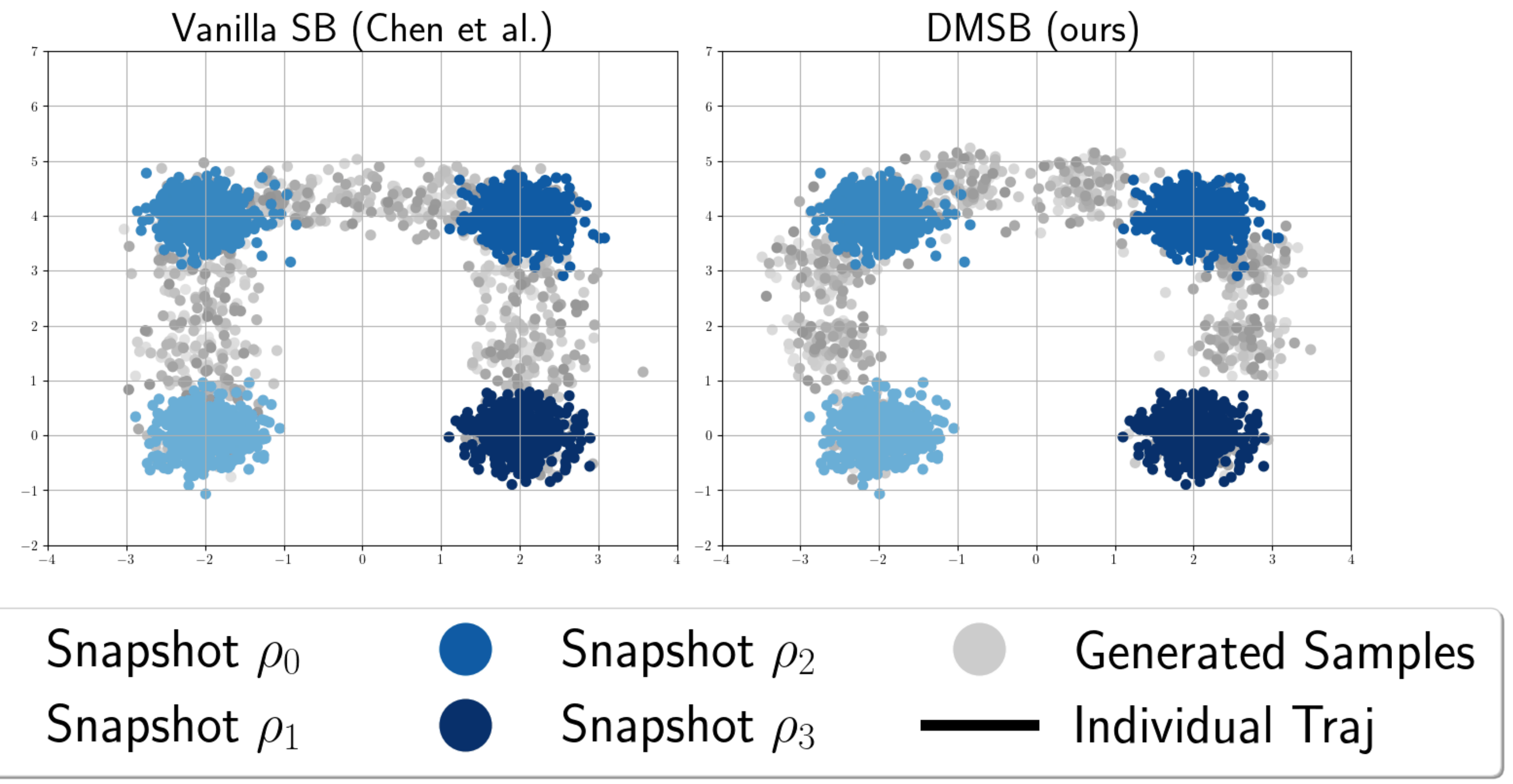
1. Some system follow optimality principle.

2. Example: least effort in the evolution of the Single Cell sequence.

3. Design Space and motivations (“Natural”)

- Newtonian Dynamics (follow Newton’s law of motion).
- Principle of least effort. (biology).
- Evolution is stochastic.

	Models	Optimality	$p_0(\cdot)$	$p_1(\cdot)$
Generative Modeling	SGM [1]	✗	$p_{\mathcal{A}}(x)$	$\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$
	CLD [2]	✗	$p_{\mathcal{A}}(x)\otimes\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$	$\mathcal{N}(\mathbf{0},\mathbf{\Sigma})\otimes\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$
Trajectory Inference	SB [3]	$\downarrow W_2 \rightarrow$ kinks	$p_{\mathcal{A}}(x)$	$p_{\mathcal{B}}(x)$
	DMSB (ours)	$\downarrow W_2 \rightarrow$ smooth	$p_{\mathcal{A}}(x)q_\theta(v\|x)$	$p_{\mathcal{B}}(x)q_\phi(v\|x)$



Momentum Schrödinger Bridge: Smooth Interpolation between distributions

[1]Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).
[2]Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion."
[3]Chen, Tianrong, Guan-Horng Liu, and Evangelos A. Theodorou. SB-FBSDE

Schrödinger Bridge Phase Space Dynamics

Two Marginals Case

Notation	Definition
$\mathbf{x}$	position variable
$\mathbf{v}$	velocity variable
$\mathbf{m}$	concatenation of $[\mathbf{x}, \mathbf{v}]^T$

Notation	Definition
ρ	position distribution $\rho(\mathbf{x})$
γ	velocity Distribution $\gamma(\mathbf{v})$
μ	distribution of $\mu(\mathbf{x}, \mathbf{v})$

$$\boxed{z^* = argmin_a \mathbb{E}\left[\int_0^1 ||z_t(x_t, v_t, t)||_2^2 dt\right]}$$

s.t

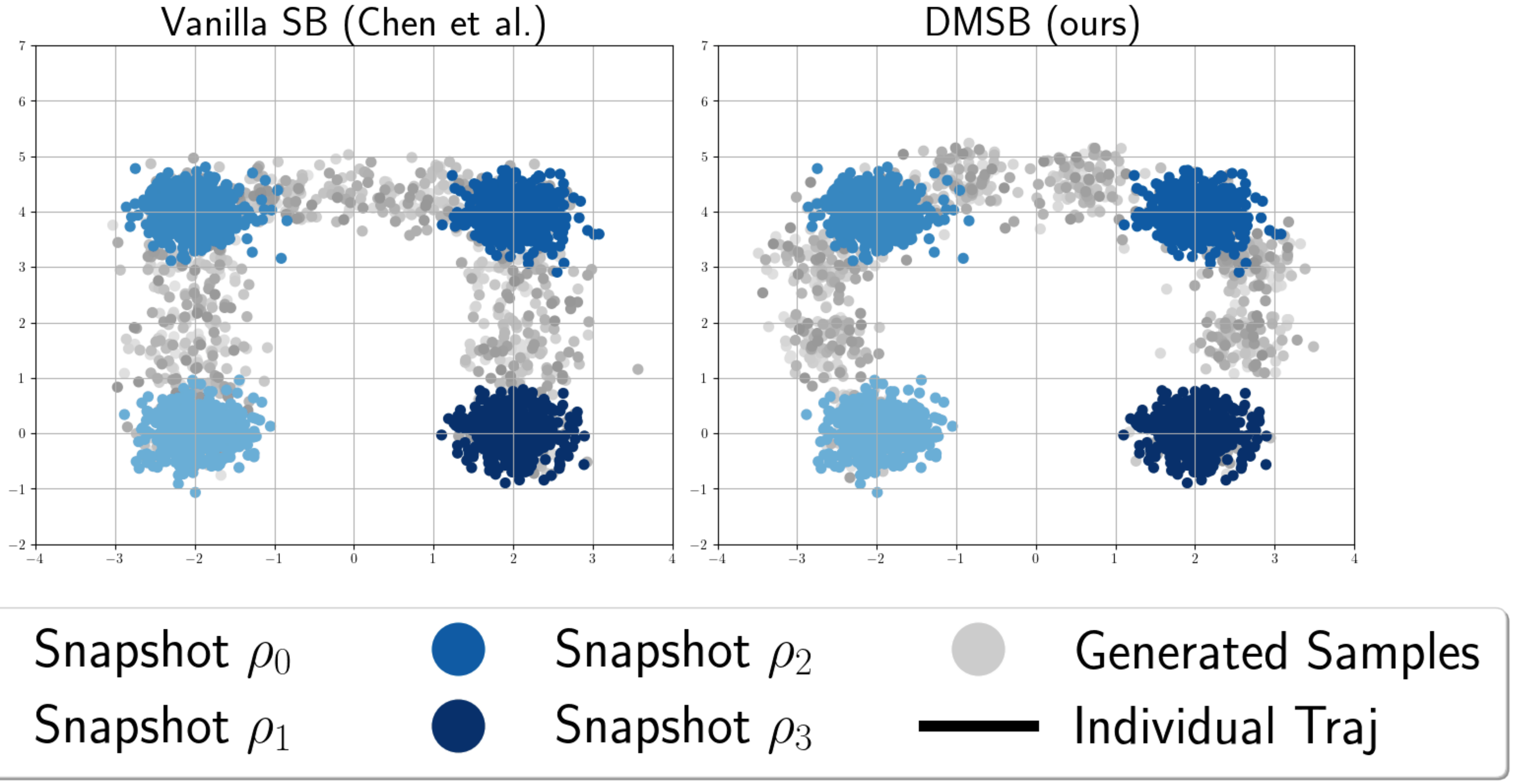
$$\boxed{\underbrace{\begin{pmatrix} \mathrm{d}\mathbf{x}_t \\ \mathrm{d}\mathbf{v}_t \end{pmatrix}}_{\mathrm{d}\mathbf{m}_t} = \underbrace{\begin{pmatrix} \mathbf{v}_t \\ \mathbf{0} \end{pmatrix}}_{\boldsymbol{f}(\mathbf{v},t)} \mathrm{d}t + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{pmatrix}}_{\mathbf{g}(t)} \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{z}_t \end{pmatrix}}_{\mathbf{Z}(t)} \mathrm{d}t + \boxed{\underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{pmatrix}}_{\mathbf{g}(t)}} \mathrm{d}\mathbf{w}_t,}$$

$$[x_0, v_0] \sim \mu_0 \qquad [x_1, v_1] \sim \mu_1$$

- $\square$ Newtonian Dynamics (follow Newton's law of motion).
- $\square$ Principle of least effort/minimize control. (Evolutionary biology).
- $\square$ Evolution is stochastic.

	SB Problem	Momentum SB Problem	Multi Marginal Momentum SB Problem
Problem Formulations	$\min_v \int_0^T \int \|\|v\|\|_2^2 dxdt$ $dx_t = v_t dt + g_t dw_t$ $x_0 \sim \rho_0, \quad x_T \sim \rho_T$	$\min_z \int_0^T \int \|\|z\|\|_2^2 dxdt$ $dx_t = v_t dt$ $dv_t = z_t dt + g dw_t$ $[x_0, v_0] \sim \rho_0 \otimes \gamma_0, \quad [x_T, v_T] \sim \rho_T \otimes \gamma_T$	$\min_z \int_0^T \int \|\|z\|\|_2^2 dxdt$ $dx_t = v_t dt$ $dv_t = z_t dt + g dw_t$ $x_i \sim \rho_i, \quad i \in [0,1,..T]$
Solutions	Baseline: Approx Likelihood Objective IPF (Chen et al. De Bortoli et al.) → Extension to Phase space →	Phase Space Approx Likelihood Objective IPF (ours) → Decouple constraints and Optimize →	DMSB (our algorithm)

Toy Example



Schrödinger Bridge Phase Space Dynamics

Two Marginals Case

$$z^* = argmin_z \mathbb{E}\left[\int_0^1 ||z_t(x_t, v_t, t)||_2^2 dt\right]$$

s.t $$\underbrace{\begin{pmatrix} \mathrm{d}\mathbf{x}_t \\ \mathrm{d}\mathbf{v}_t \end{pmatrix}}_{\mathrm{d}\mathbf{m}_t} = \underbrace{\begin{pmatrix} \mathbf{v}_t \\ \mathbf{0} \end{pmatrix}}_{\boldsymbol{f}(\mathbf{v},t)} \mathrm{d}t + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{pmatrix}}_{\mathbf{g}(t)} \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{z}_t \end{pmatrix}}_{\mathbf{Z}(t)} \mathrm{d}t + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{pmatrix}}_{\mathbf{g}(t)} \mathrm{d}\mathbf{w}_t,$$

$$[x_0, v_0] \sim \mu_0 \qquad [x_1, v_1] \sim \mu_1$$

We want to use Bregman Iteration

$$\min_{\pi \in \Pi(\mu_0, \mu_T)} KL(\pi|\xi) \quad s.t \quad \pi = \mathrm{Law}(\mathbf{x}, \mathbf{v}) : \underbrace{\begin{pmatrix} \mathrm{d}\mathbf{x}_t \\ \mathrm{d}\mathbf{v}_t \end{pmatrix}}_{\mathrm{d}\mathbf{m}_t} = \underbrace{\begin{pmatrix} \mathbf{v}_t \\ \mathbf{0} \end{pmatrix}}_{\boldsymbol{f}(\mathbf{v},t)} \mathrm{d}t + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{pmatrix}}_{\mathbf{g}(t)} \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{z}_t \end{pmatrix}}_{\mathbf{Z}(t)} \mathrm{d}t + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g_t \end{pmatrix}}_{\mathbf{g}(t)} \mathrm{d}\mathbf{w}_t,$$

Bregman Iteration

1. Problem Formulation: $\inf_{\boxed{\pi \in K}} KL(\pi \,|\, \boxed{\xi})$

Reference Path measure.
Typical choice: Wiener Process.

Constraint Sets.

1.Method: Given constraint sets: $K = \cap_0^L K_l$

Bregman Projection

dth **Bregman Iteration**

$$\inf_{\pi \in K_0} KL(\pi \,|\, \xi) \xrightarrow{\pi^{(d,0)}} \inf_{\pi \in K_1} KL(\pi \,|\, \pi^{(d,0)}) \xrightarrow{\pi^{(d,1)}} \inf_{\pi \in K_2} KL(\pi \,|\, \pi^{(d,1)}) \xrightarrow{\pi^{(d,2)}}$$

$(d+1)$th **Bregman Iteration**

$$\inf_{\pi \in K_0} KL(\pi \,|\, \pi^{(d,2)}) \xrightarrow{\pi^{(d+1,0)}} \inf_{\pi \in K_1} KL(\pi \,|\, \pi^{(d+1,0)}) \xrightarrow{\pi^{(d+1,1)}} \inf_{\pi \in K_2} KL(\pi \,|\, \pi^{(d+1,1)})$$

Schrödinger Bridge Phase Space Dynamics

$$\min_{\pi\in\Pi(\mu_0,\mu_T)} KL(\pi|\xi) \quad s.t \quad \pi = \mathrm{Law}(\mathbf{x},\mathbf{v}) : \underbrace{\begin{pmatrix}\mathrm{d}\mathbf{x}_t\\ \mathrm{d}\mathbf{v}_t\end{pmatrix}}_{\mathrm{d}\mathbf{m}_t} = \underbrace{\begin{pmatrix}\mathbf{v}_t\\ \mathbf{0}\end{pmatrix}}_{\boldsymbol{f}(\mathbf{v},t)} \mathrm{d}t + \underbrace{\begin{pmatrix}0 & 0\\ 0 & g_t\end{pmatrix}}_{\mathbf{g}(t)} \underbrace{\begin{pmatrix}0\\ \mathbf{z}_t\end{pmatrix}}_{\mathbf{Z}(t)} \mathrm{d}t + \underbrace{\begin{pmatrix}0 & 0\\ 0 & g_t\end{pmatrix}}_{\mathbf{g}(t)} \mathrm{d}\mathbf{w}_t,$$

$$K_0 = \Pi(\mu_0, \cdot\,) = \left\{ \pi \,\middle|\, \mu_0 = \int_T^0 \pi_{0,1} dx_T \right\} \qquad K_1 = \Pi(\,\cdot\,, \mu_T) = \left\{ \pi \,\middle|\, \mu_T = \int_0^T \pi_{0,1} dx_0 \right\}$$

Bregman Projection

dth **Bregman Iteration**

$\inf_{\pi \in K_0} KL(\pi \,|\, \xi) \xrightarrow{\pi^{(d,0)}} \inf_{\pi \in K_1} KL(\pi \,|\, \pi^{(d,0)}) \xrightarrow{\pi^{(d,1)}} \inf_{\pi \in K_2} KL(\pi \,|\, \pi^{(d,1)}) \xrightarrow{\pi^{(d,2)}}$

$(d+1)$th **Bregman Iteration**

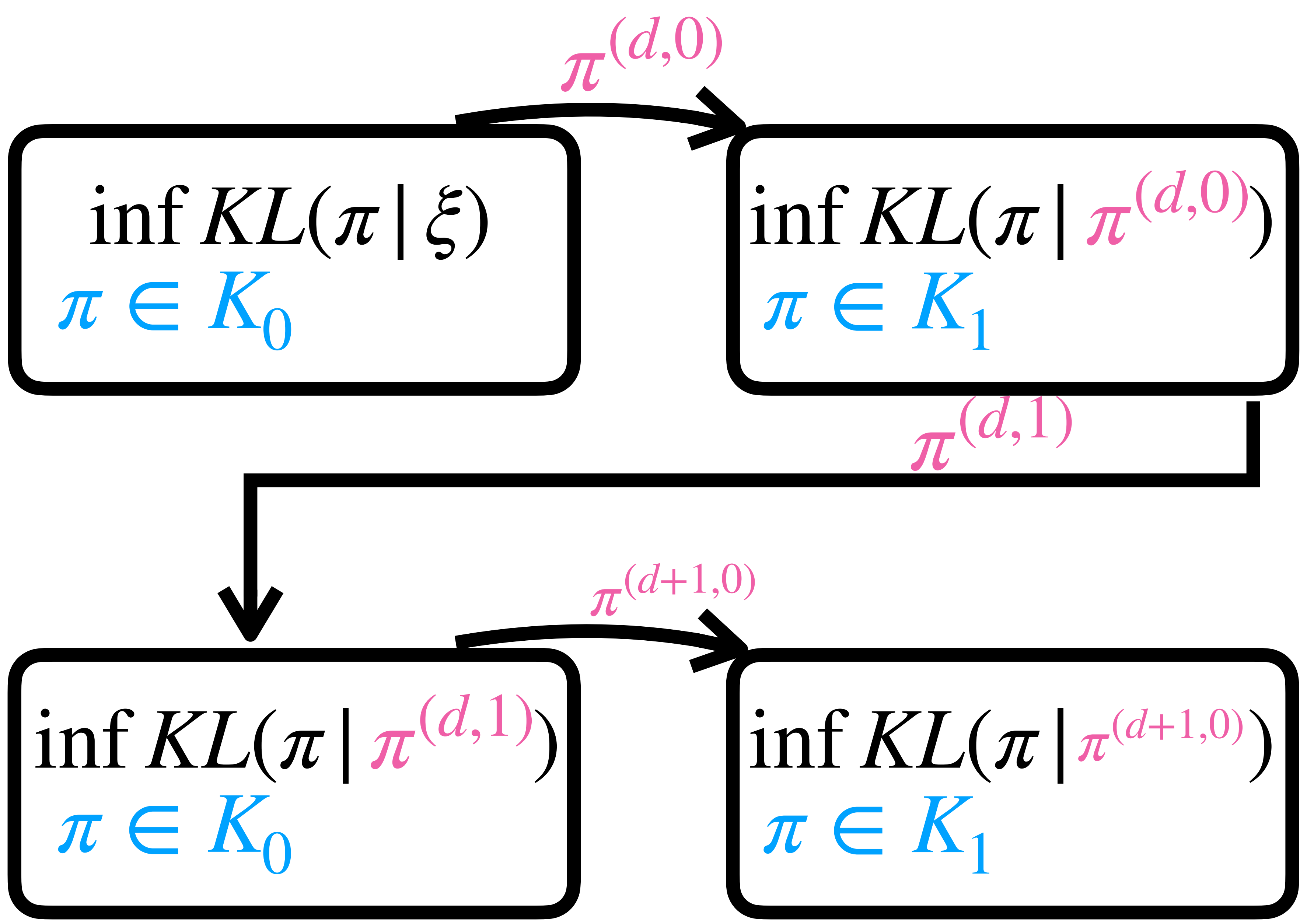
$\inf_{\pi \in K_0} KL(\pi \,|\, \pi^{(d,2)}) \xrightarrow{\pi^{(d+1,0)}} \inf_{\pi \in K_1} KL(\pi \,|\, \pi^{(d+1,0)}) \xrightarrow{\pi^{(d+1,1)}} \inf_{\pi \in K_2} KL(\pi \,|\, \pi^{(d+1,1)})$

Schrödinger Bridge Phase Space Dynamics

$$\min_{\pi\in\Pi(\mu_0,\mu_T)} KL(\pi|\xi) \quad s.t \quad \pi=\mathrm{Law}(\mathbf{x},\mathbf{v}):\underbrace{\begin{pmatrix}\mathrm{d}\mathbf{x}_t\\ \mathrm{d}\mathbf{v}_t\end{pmatrix}}_{\mathbf{dm}_t}=\underbrace{\begin{pmatrix}\mathbf{v}_t\\ \mathbf{0}\end{pmatrix}}_{\boldsymbol{f}(\mathbf{v},t)}\mathrm{d}t+\underbrace{\begin{pmatrix}\mathbf{0}&\mathbf{0}\\ \mathbf{0}&g_t\end{pmatrix}}_{\mathbf{g}(t)}\underbrace{\begin{pmatrix}\mathbf{0}\\ \mathbf{z}_t\end{pmatrix}}_{\mathbf{Z}(t)}\mathrm{d}t+\underbrace{\begin{pmatrix}\mathbf{0}&\mathbf{0}\\ \mathbf{0}&g_t\end{pmatrix}}_{\mathbf{g}(t)}\mathrm{d}\mathbf{w}_t,$$

dth **Bregman Iteration**

$(d+1)$th **Bregman Iteration**



Schrödinger Bridge Phase Space Dynamics

Forward: $\pi^\theta : \quad dm_t = [f + g_t Z_t^\theta]dt + g_t dw_t$

Backward: $\pi^\phi : \quad dm_t = [f + g_t \hat{Z}_t^\phi]dt + g_t dw_t$

$$K_0 = \Pi(\mu_0, \cdot\,) = \left\{ \pi \,|\, \mu_0 = \int_T^0 \pi_{0,1} dx_T \right\} \qquad K_1 = \Pi(\,\cdot\,, \mu_T) = \left\{ \pi \,|\, \mu_T = \int_0^T \pi_{0,1} dx_0 \right\}$$

$$\begin{tikzcd} \boxed{\begin{array}{l} \inf KL(\pi^\theta \,|\, \pi^\phi) \\ \pi^\theta \in K_0 \end{array}} \arrow[r, bend left, "\pi^\theta"] & \boxed{\begin{array}{l} \inf KL(\pi^\phi \,|\, \pi^\theta) \\ \pi^\phi \in K_1 \end{array}} \arrow[l, bend left, "\pi^\phi"] \end{tikzcd} \xrightarrow{\text{Converge}} \min_{\pi \in \Pi(\mu_0, \mu_1)} KL(\pi \,|\, \xi)$$

Iterative Proportional Fitting (IPF)

(Special case of Bregman Iteration)

Schrödinger Bridge Phase Space Dynamics

Forward: π^θ : $dm_t = [f + g_t Z_t^\theta]dt + g_t dw_t$

Backward: π^ϕ : $dm_t = [f + g_t \hat{Z}_t^\phi]dt + g_t dw_t$

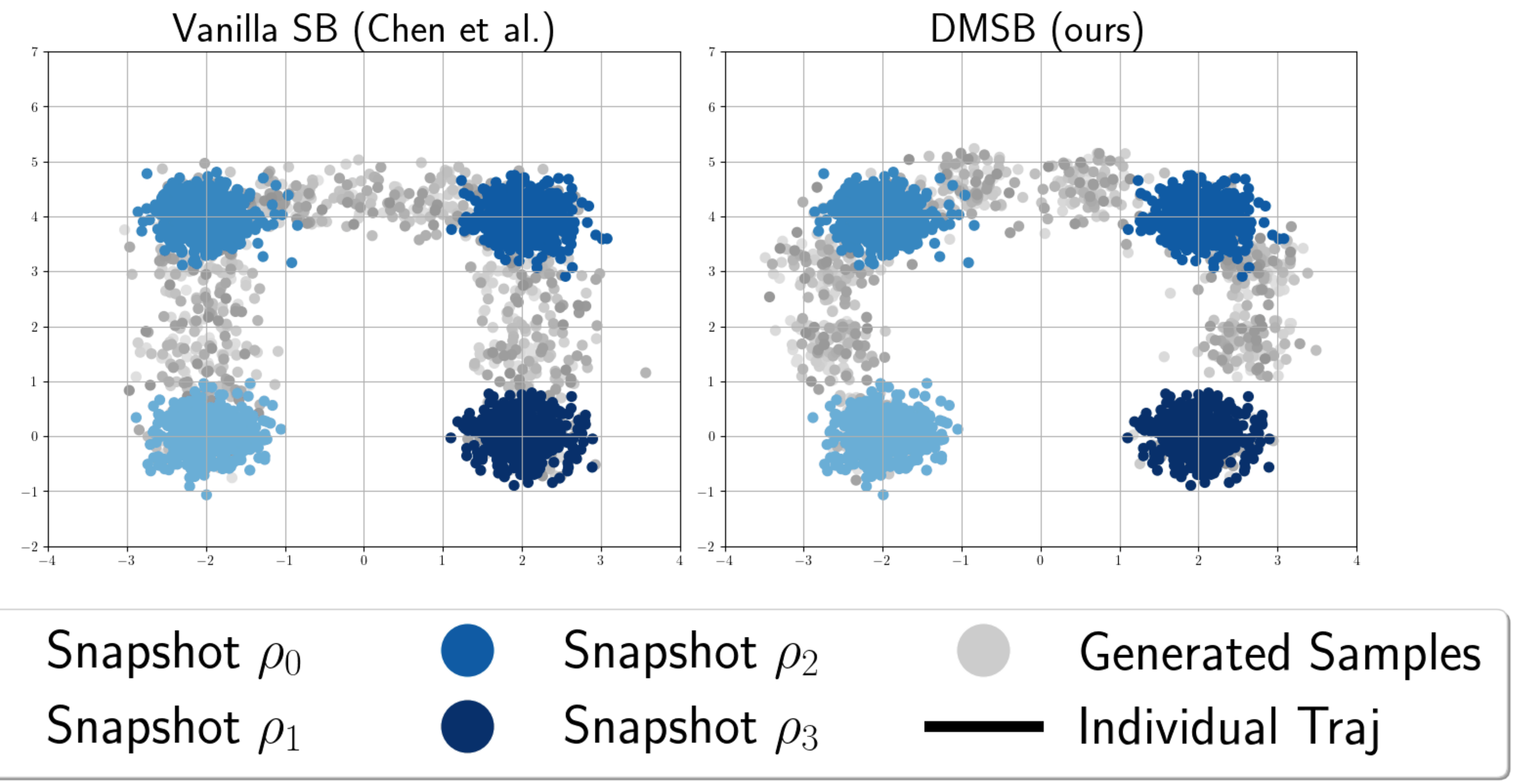
$K_0 = \Pi(\mu_0, \cdot\,) = \left\{ \pi \,|\, \mu_0 = \int_T^0 \pi_{0,1} dx_T \right\}$ $\quad$ $K_1 = \Pi(\,\cdot\,, \mu_T) = \left\{ \pi \,|\, \mu_T = \int_0^T \pi_{0,1} dx_0 \right\}$

$$\boxed{\begin{matrix}\inf KL(\pi^\theta | \ \pi^\phi \ ) \\ \pi^\theta \in K_0\end{matrix}} \propto \min_\theta \int_0^T \mathbb{E}_{\pi^\phi} || z_t^\theta + \hat{z}_t^\phi - g \nabla_v \log p_t^\phi ||_2^2 dt$$

$$\propto \min_\theta \int_0^T \mathbb{E}_{\pi^\phi} \left\{ || z_t^\theta ||_2^2 + 2\langle z_t^\theta, \hat{z}_t^\phi \rangle + \nabla_v (g z_t^\theta) \right\} dt$$

	SB Problem	Momentum SB Problem	Multi Marginal Momentum SB Problem
Problem Formulations	$\min_v \int_0^T \int \lvert\lvert v \rvert\rvert_2^2 dxdt$ $dx_t = v_t dt + g_t dw_t$ $x_0 \sim \rho_0, \quad x_T \sim \rho_T$	$\min_z \int_0^T \int \lvert\lvert z \rvert\rvert_2^2 dxdt$ $dx_t = v_t dt$ $dv_t = z_t dt + g dw_t$ $[x_0, v_0] \sim \rho_0 \otimes \gamma_0, \quad [x_T, v_T] \sim \rho_T \otimes \gamma_T$	$\min_z \int_0^T \int \lvert\lvert z \rvert\rvert_2^2 dxdt$ $dx_t = v_t dt$ $dv_t = z_t dt + g dw_t$ $x_i \sim \rho_i, \quad i \in [0,1,..T]$
Solutions	Baseline: Approx Likelihood Objective IPF (Chen et al. De Bortoli et al.) → Extension to Phase space →	Phase Space Approx Likelihood Objective IPF (ours) → Decouple constraints and Optimize →	DMSB (our algorithm)

Toy Example



Schrödinger Bridge Phase Space Dynamics

Multi-Marginal case

We **may** solve multi-marginal momentum Schrödinger Bridge by **Bregman Iteration.**

Two problems to solve:

1. Is Bregman Projection easy to conduct?
2. How can we get velocity?

Schrödinger Bridge Phase Space Dynamics

Multi-Marginal case

We **may** solve multi-marginal momentum Schrödinger Bridge by **Bregman Iteration.**

Two problems to solve:

1. Is Bregman Projection easy to conduct?
2. How can we get velocity?

Schrödinger Bridge Phase Space Dynamics

Constraints in multi-marginal case

Proposition 4.1 ([1]). *The dynamical mmmSB with multiple marginal constraints reads:*

$$\min_{\pi} \mathcal{J}(\pi) := \sum_{i=0}^{N-1} KL\left(\pi_{t_i:t_{i+1}} | \xi_{t_i:t_{i+1}}\right), \quad s.t \quad \pi \in \mathcal{K} := \cap_{i=0}^{N} \mathcal{K}_{t_i}$$

where:

$$\mathcal{K}_{t_0} = \left\{ \int \pi_{t_0:t_1} \mathbf{dm}_{t_1} = \mu_{t_0}, \int \mu_{t_0} \mathbf{dv}_{t_0} = \rho_{t_0} \right\}$$

$$\mathcal{K}_{t_N} = \left\{ \int \pi_{t_{N-1}:t_N} \mathbf{dm}_{t_{N-1}} = \mu_{t_N}, \int \mu_{t_N} \mathbf{dv}_{t_N} = \rho_{t_N} \right\}$$

$$\mathcal{K}_{t_i} = \left\{ \int \pi_{t_i:t_{i+1}} \mathbf{dm}_{t_{i+1}} = \mu_{t_i}, \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} = \mu_{t_i}, \int \mu_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\},$$

and $\mathcal{K}$ is the intersection of close convex set of $\mathcal{K}_{t_i}$.

Similar to previous constraint set

Feasible

Cannot be directly optimized by NN. [1]

Notation Recap:

Notation	Definition
$\mathbf{x}$	position variable
$\mathbf{v}$	velocity variable
$\mathbf{m}$	concatenation of $[\mathbf{x}, \mathbf{v}]^{\mathsf{T}}$

Notation	Definition
ρ	position distribution $\rho(\mathbf{x})$
γ	velocity Distribution $\gamma(\mathbf{v})$
μ	distribution of $\mu(\mathbf{x}, \mathbf{v})$

[1] Chen, Yongxin, et al. "Multi-marginal Schrödinger bridges." *International Conference on Geometric Science of Information*. Cham: Springer International Publishing, 2019.

Schrödinger Bridge Phase Space Dynamics

Bregman Projection in multi-marginal case

Proposition 4.1 ([1]). *The dynamical mmmSB with multiple marginal constraints reads:*

$$\min_{\pi} \mathcal{J}(\pi) := \sum_{i=0}^{N-1} KL\left(\pi_{t_i:t_{i+1}} | \xi_{t_i:t_{i+1}}\right), \quad s.t \quad \pi \in \mathcal{K} := \cap_{i=0}^{N} \mathcal{K}_{t_i}$$

where:

$$\mathcal{K}_{t_0} = \left\{ \int \pi_{t_0:t_1} \mathbf{dm}_{t_1} = \mu_{t_0}, \int \mu_{t_0} \mathbf{dv}_{t_0} = \rho_{t_0} \right\}$$

$$\mathcal{K}_{t_N} = \left\{ \int \pi_{t_{N-1}:t_N} \mathbf{dm}_{t_{N-1}} = \mu_{t_N}, \int \mu_{t_N} \mathbf{dv}_{t_N} = \rho_{t_N} \right\}$$

$$\mathcal{K}_{t_i} = \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \mu_{t_i}, \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} = \mu_{t_i}, \int \mu_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\},$$

and $\mathcal{K}$ is the intersection of close convex set of $\mathcal{K}_{t_i}$.

Similar to previous constraint set

Feasible

Can we decompose the constraint sets?

[1] Chen, Yongxin, et al. "Multi-marginal Schrödinger bridges." *International Conference on Geometric Science of Information*. Cham: Springer International Publishing, 2019.

Schrödinger Bridge Phase Space Dynamics

Bregman Projection in multi-marginal case

Proposition 4.1 ([1]). *The dynamical mmmSB with multiple marginal constraints reads:*

$$\min_{\pi} \mathcal{J}(\pi) := \sum_{i=0}^{N-1} KL\left(\pi_{t_i:t_{i+1}} | \xi_{t_i:t_{i+1}}\right), \quad s.t \quad \pi \in \mathcal{K} := \cap_{i=0}^{N} \mathcal{K}_{t_i}$$

where:

$$\mathcal{K}_{t_0} = \left\{ \int \pi_{t_0:t_1} \mathbf{dm}_{t_1} = \mu_{t_0}, \int \mu_{t_0} \mathbf{dv}_{t_0} = \rho_{t_0} \right\}$$

$$\mathcal{K}_{t_N} = \left\{ \int \pi_{t_{N-1}:t_N} \mathbf{dm}_{t_{N-1}} = \mu_{t_N}, \int \mu_{t_N} \mathbf{dv}_{t_N} = \rho_{t_N} \right\}$$

$$\mathcal{K}_{t_i} = \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \mu_{t_i}, \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} = \mu_{t_i}, \int \mu_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\},$$

and $\mathcal{K}$ is the intersection of close convex set of $\mathcal{K}_{t_i}$.

$$\mathcal{K}_{t_i} = \cap_{r=0}^{2} \mathcal{K}_{t_i}^{r}, \quad \text{where} \quad \begin{aligned} \mathcal{K}_{t_i}^{0} &= \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \hat{\mu}_{t_i}, \int \hat{\mu}_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\} \\ \mathcal{K}_{t_i}^{1} &= \left\{ \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} = \mu_{t_i}, \int \mu_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\} \\ \mathcal{K}_{t_i}^{2} &= \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} \right\}. \end{aligned}$$

K

$K_{boundary}$

$K_{bridge} := \cap_{i=1}^{N-1} K_{t_i}^{2}$

Schrödinger Bridge Phase Space Dynamics

Bregman Projection in multi-marginal case

Proposition 4.1 ([1]). *The dynamical mmmSB with multiple marginal constraints reads:*

$$\min_{\pi} \mathcal{J}(\pi) := \sum_{i=0}^{N-1} KL\left(\pi_{t_i:t_{i+1}} | \xi_{t_i:t_{i+1}}\right), \quad s.t \quad \pi \in \mathcal{K} := \cap_{i=0}^{N} \mathcal{K}_{t_i}$$

where:

$$\mathcal{K}_{t_0} = \left\{ \int \pi_{t_0:t_1} \mathbf{dm}_{t_1} = \mu_{t_0}, \int \mu_{t_0} \mathbf{dv}_{t_0} = \rho_{t_0} \right\}$$

$$\mathcal{K}_{t_N} = \left\{ \int \pi_{t_{N-1}:t_N} \mathbf{dm}_{t_{N-1}} = \mu_{t_N}, \int \mu_{t_N} \mathbf{dv}_{t_N} = \rho_{t_N} \right\}$$

$$\mathcal{K}_{t_i} = \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \mu_{t_i}, \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} = \mu_{t_i}, \int \mu_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\},$$

and $\mathcal{K}$ is the intersection of close convex set of $\mathcal{K}_{t_i}$.

$$\mathcal{K}_{t_i} = \cap_{r=0}^{2} \mathcal{K}_{t_i}^{r}, \quad \text{where} \quad \begin{aligned} \mathcal{K}_{t_i}^{0} &= \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \hat{\mu}_{t_i}, \int \hat{\mu}_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\} \\ \mathcal{K}_{t_i}^{1} &= \left\{ \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} = \mu_{t_i}, \int \mu_{t_i} \mathbf{dv}_{t_i} = \rho_{t_i} \right\} \\ \mathcal{K}_{t_i}^{2} &= \left\{ \int \pi_{t_i:t_{t+1}} \mathbf{dm}_{t_{i+1}} = \int \pi_{t_{i-1}:t_i} \mathbf{dm}_{t_{i-1}} \right\}. \end{aligned}$$

K

$K_{boundary}$

$K_{bridge} := \cap_{i=1}^{N-1} K_{t_i}^{2}$

Just Continuous path measure

Schrödinger Bridge Phase Space Dynamics

Multi-Marginal case

We can solve multi-marginal momentum Schrödinger Bridge by **Bergman Iteration.**

Two problems to solve:

1. Bregman Projection easy to conduct?
2. How can we get velocity?

Schrödinger Bridge Phase Space Dynamics

Sample Velocity

Recap:

$$\boxed{\begin{array}{l}\inf KL(\pi^\theta \,|\, \pi^\phi \;) \\ \pi^\theta \in K_0\end{array}}$$

$$\propto \min_\theta \int_0^T \mathbb{E}_{\pi^\phi} || z_t^\theta + \hat{z}_t^\phi - g \nabla_v \log p_t^\phi ||_2^2 dt$$

$$\propto \min_\theta \int_0^T \mathbb{E}_{\pi^\phi} \left\{ || z_t^\theta ||_2^2 + 2\langle z_t^\theta, \hat{z}_t^\phi \rangle + \nabla_v (g z_t^\theta) \right\} dt$$

Once can sample velocity by Langevin dynamics induced by estimated score function parameterized by $(z_t^\theta + \hat{z}_t^\phi)/g$.

Schrödinger Bridge Phase Space Dynamics

Multi-Marginal case

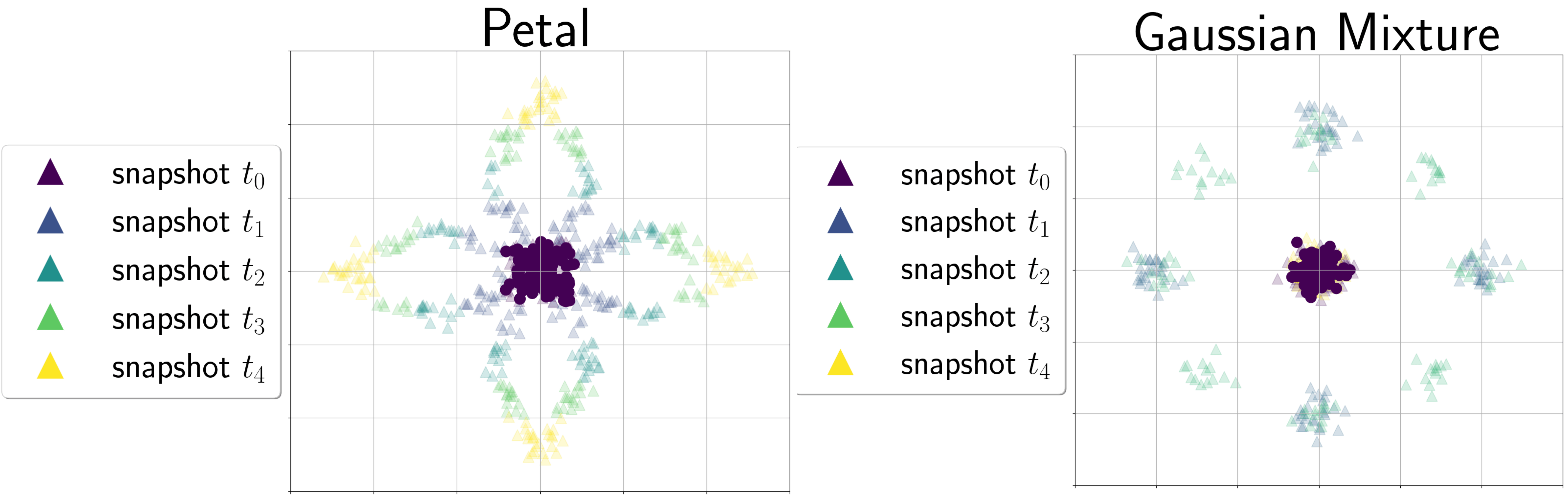
We can solve multi-marginal momentum Schrödinger Bridge by **Bregman Iteration.**

Two problems to solve:

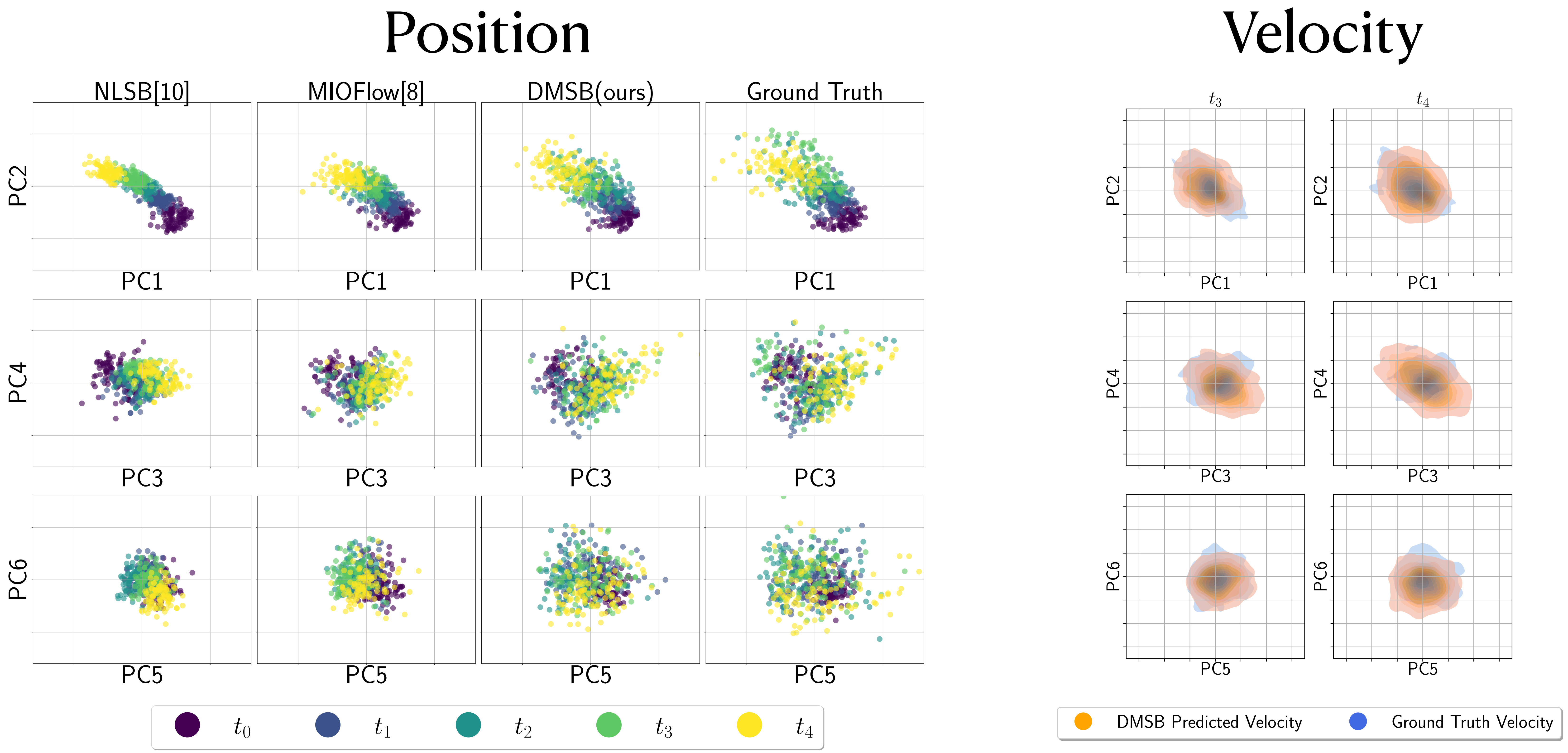
1. ~~Bregman Projection easy to conduct?~~
2. ~~How can we get velocity?~~

2D toy dataset:

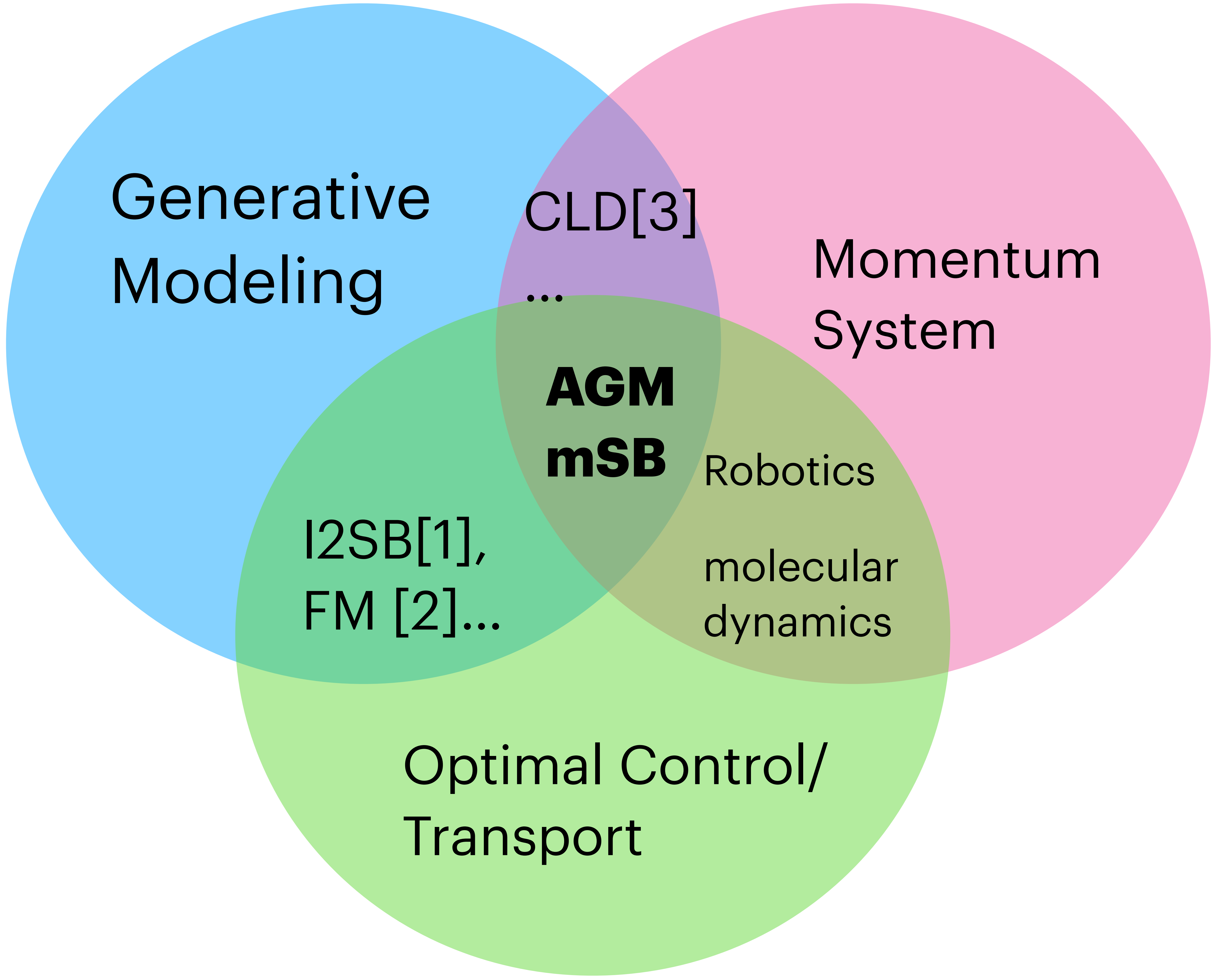
(Petal and Gaussian Mixture)



100D RNA sequence dataset



Summary



AGM:

1. **Momentum System:** extra information for sampling-hop.
2. **Optimal control:**
    1. straight trajectory.
    2. Exponential Integrator

mSB:

1. **Momentum System:**
    1. Newtonian dynamics.
2. **Optimal Transport:**
    1. smooth path measure.

[1] Liu, Guan-Horng, et al. "I $^ 2$ SB: Image-to-Image Schr\" odinger Bridge." *arXiv preprint arXiv:2302.05872* (2023).
[2] Lipman, Yaron, et al. "Flow matching for generative modeling." *arXiv preprint arXiv:2210.02747* (2022).
[3] Dockhorn, Tim, Arash Vahdat, and Karsten Kreis. "Score-based generative modeling with critically-damped langevin diffusion."

Thanks to my awesome co-authors

Acceleration Generative Model


Jiatao Gu


Laurent Dinh


Evangelos Theodorou


Joshua Susskind


Shuangfei Zhai

Momentum Schrödinger Bridge


Guan-Horng Liu


Molei Tao


Evangelos Theodorou

Q&A