Learning diffusion at lightspeed

 $,$ (1)

ETHzürich

 $\text{effects mismatch} = W_2(\mu_1, \hat{\mu}_1)^2$

\blacksquare diffusion, we want to find the parameters energy function \mathcal{L} of the parameters energy function \mathcal{L}

2.1 Preliminaries Intuition in \mathbb{R}^d

The gradient of ^ϑ : ^R^d [↑] ^R is [→]^ϑ [↓] ^R^d and the Jacobian of ^ϖ : ^R^d [↑] ^Rⁿ is [→]^ϖ [↓] ^Rⁿ↓^d. We Consider the analog of (1) in the \mathbb{R}^d ,

• Diffusion processes are widespread in many natural processes. observed the predicted trajectory and the predicted trajectory obtained in the Junction of Typically characterized by three terms: a drift term due to a notential • Typically characterized by three terms: a drift term due to a potential field, the interaction with other particles, and a stochastic term. • Existing methods either rely on having access to particles trajectories in discribution of the state of the state of the state of the next state of the next state in the next state in the next state $\frac{1}{100}$ or are data and compute inefficient (often not possible) or are data and compute inefficient. field, the interaction with other particles, and a stochastic term. (often not possible) or are data and compute inefficient.

Tho IKO cchamo s are $\frac{1}{2}$ in the diffusion process is what $\frac{1}{2}$ minimizes. The diffusion process is what $\frac{1}{2}$ The JKO scheme

 $\mathcal{S}(\mathcal{S})$. $\mathcal{S}(\mathcal{S})$. 28]. $\mathcal{S}(\mathcal{S})$. $W(t)$ the Wiener process. It turns out the resulting particles trajectory α optimization problem in the problem in the problem in the exploited via the exploited via the exploited via the explo to learn the energy functional governing the underlying diffusion process from population process from population
The underlying diffusion process from population data, and population data, and population data, and in the where $X(t)$ is the state of the particle, $V(x)$ the driving potential, and can be described via the JKO scheme:

NaN JKOnet [∗] JKOnet ∗ V JKOnet ∗ $\hbox{}_{\!\! L}^*$ $\!\circ$ JKOnet $\hbox{}_{\!\! L}^*$

$$
dX(t) = -\boldsymbol{\nabla} V(X(t))dt + \sqrt{2\beta}dW(t),
$$

2 Diffusion processes via optimal transport

$$
\mu_{t+1} = \operatorname*{argmin}_{\mu \in \mathcal{P}(\mathbb{R}^d)} J(\mu) + \frac{1}{2\tau} W_2(\mu, \mu_t)^2, \tag{1}
$$

where J is an energy functional and $\tau > 0$ is the time discretization. $\hspace{0.2cm}$ \sim A general energy functional: allows us to allow us to all the state of the state of the state of the s

> • Observability of the different energy terms; • Can optimality conditions in $\mathcal{P}(\mathbb{R}^d)$ be helpful

th 2024 – December 15th 2024

$$
J(\mu) = \underbrace{\int_{\mathbb{R}^d} V(x) \mathrm{d}\mu(x)}_{\text{potential}} + \underbrace{\int_{\mathbb{R}^d \times \mathbb{R}^d} U(x - y) \mathrm{d}(\mu \times \mu)(x, y)}_{\text{interaction}} + \underbrace{\beta \int_{\mathbb{R}^d} \rho(x) \log(\rho(x)) \mathrm{d}x}_{\text{internal}}.
$$

The Fokker-Planck equation,

 $\mathcal{O}(\mu(\iota, x)) = \nabla \cdot (\nabla \mu(\kappa) \cdot (1 + \kappa) + 8 \nabla^2 \cdot (1 + \kappa)$ $p_t = \mathbf{v} \cdot (\mathbf{v} \mathbf{v}(\lambda) p(\iota, \lambda)) + \rho \mathbf{v} \mathbf{v}(\iota, \lambda),$ σ $\partial \rho(t,x)$ ∂t $= \boldsymbol{\nabla} \boldsymbol{\cdot} (\boldsymbol{\nabla} V(\textit{\textbf{x}}) \rho(t, \textit{\textbf{x}})) + \beta \nabla^2 \rho(t, \textit{\textbf{x}}),$

describes the thire evolution of the distribution ρ or a set of particles undergoing drift and diffusion, the observed transport of the observed transport of the observed transport of describes the time evolution of the distribution ρ of a set of particles

$$
\mathsf{x}_{t+1}=\mathop{\rm argmin}_{\mathsf{x}\in\mathbb{R}^d}J(\mathsf{x})+\frac{1}{2\tau}\|\mathsf{x}-\mathsf{x}_t\|^2.
$$

We replace the above by its first-order optimality condition

internal

$$
\nabla J(x_{t+1}) + \frac{1}{\tau}(x_{t+1} - x_t) = 0.
$$

Given a dataset $(x_0, x_1, \ldots, x_{\tau})$, we find J as:

$$
\min_{j} \sum_{t=0}^{T-1} \left\| \nabla J(x_{t+1}) + \frac{1}{\tau} (x_{t+1} - x_t) \right\|^2.
$$

Our work enables an analogous result in the Wasserstein space.

38th Conference on Neural Information Processing Systems (NeurIPS 2024, oral)

First-order optimality conditions for the JKO scheme

Our analysis suggest that given a populations dataset $(\mu_0, \mu_1, \dots, \mu_T)$ we seek the parameters θ minimizing

$$
\sum_{t=0}^{T-1}\int_{\mathbb{R}^d\times\mathbb{R}^d}\left\|\boldsymbol{\nabla}V_{\theta}(x_{t+1})+\int_{\mathbb{R}^d}\boldsymbol{\nabla}U_{\theta}(x_{t+1}-x_{t+1}')\mathrm{d}\mu_{t+1}(x_{t+1}')
$$
\n
$$
+\beta_{\theta}\frac{\boldsymbol{\nabla} \rho_{t+1}(x_{t+1})}{\rho_{t+1}(x_{t+1})}+\frac{1}{\tau}(x_{t+1}-x_t)\right\|^2\mathrm{d}\gamma_t(x_t,x_{t+1}),
$$
 (2)

where $\gamma_0, \gamma_1, \ldots, \gamma_T$ are optimal transport plans between μ_t and μ_{t+1} .

- γ_t can be computed once, before-hand, and efficiently;
- when J_{θ} is a neural network, we minimize (2) via gradient descent;
- when J_{θ} is parametrized linearly, we have a closed-form solution.

Training at lightspeed

Different potential $V(x)$ driving the diffusion.

l,V JKOnet JKOnet-vanilla

Eye-candies

Level curves of different potentials (green: real; blue: estimated).

Antonio Terpin, Nicolas Lanzetti, Martin Gadea, Florian Dörfler

Scaling to high-dimensions

- Sub-linear error growth.
- Negligible increase in training time.

N. of particles

BO

 $\overline{\sigma}$

 \overline{z}

 \overline{L}

 \geq

Learning general energy terms

Additional error sources: • Sampling error (internal energy) • Estimation of the densities Nonetheless: our method (JKOnet*, non-linear and linearly parametrized) recovers all the energy terms. Open question: observability of the different energy terms?

Learning single-cell diffusion dynamics

• Intuition: cellular evolution minimizes (some) energy; • We can account for unobserved variables via time-varying energies; of the computational cost of the other methods: less than minute vs hours.

-
-
- State-of-the-art accuracy at a fraction
-

Cool stuff worth looking at next

• Fast distillation of diffusion models for one- or few-steps generation;

-
-
- in your work? Reach out!

