Partially Observed Trajectory Inference using Optimal Transport and a Dynamics Prior

Anming Gu [1] Edward Chien [1] Kristjan Greenewald [2]

[1]Boston University [2]IBM Research

Latent Trajectory Inference

Trajectory inference seeks to recover the temporal dynamics of a population from snapshots of its (uncoupled) temporal marginals, i.e. where observed particles are not tracked over time. Prior works [1, 2] framed the problem under a stochastic differential equation (SDE) model in observation space, proposed a non-parametric maximum likelihood estimator, and utilized the mean-field Langevin algorithm to solve this problem. We consider extending this setting where the dynamics is not directly observed using observable state-space models.

Contributions

We introduce the partially observed trajectory problem using observable state-space models and a dynamics prior. We extend the the estimator and algorithms from [1, 2] to our setting and show:

- **Asymptotic consistency** guarantees to underlying SDE as number of observations becomes dense.
- **Convergence** of algorithm.

Problem Setup

Let $\mathcal{X}$ be the latent space and $\mathcal{Y}$ be the observation space. We have a population following these dynamics on $\mathcal{X}$ over the time interval $[0, 1]$ with initial condition $\mathbf{P}_0$:

$$dX_t = -\Xi(t, X_t)dt - \nabla\Psi(t, X_t)dt + \sqrt{\tau}dB_t, \qquad X_0 \sim \mathbf{P}_0. \tag{1}$$

- B_t is a Brownian motion
- τ is the *known* diffusivity,
- $\Xi : C([0, 1] \times \mathcal{X} : \mathcal{X})$ is a *known* divergence-free vector field
- $\Psi : C^2([0, 1] \times \mathcal{X})$ is an *unknown* potential function

Let $g : \mathcal{X} \to \mathcal{Y}$ be an observation function.

We have T observation times with $0 \leq t_1^T < \cdots \leq t_T^T \leq 1$ with N_i^T i.i.d. datapoints at each time $i \in [T]$:

$$\left\{Y_{i,j}^T\right\}_{j=1}^{N_i^T} \sim g_\sharp \mathbf{P}_{t_i^T},$$

where $\mathbf{P}_t$ is the marginal of Eq. (1) at time t and $g_\sharp \mathbf{P}_t$ is its pushforward onto observation space $\mathcal{Y}$. These datapoints form the empirical distribution:

$$\hat{\rho}_i^T = \frac{1}{N_i^T} \sum_{j=1}^{N_i^T} \delta_{Y_{i,j}^T}$$

Key assumption: observability

Assume Ψ is unknown but restricted to a class $\mathcal{C}_\Psi$. We say the tuple $(g, \Xi, \mathcal{C}_\Psi)$ is $\mathcal{C}_\Psi$-ensemble observable if, given g, Ξ, τ and marginals $g_\sharp \mathbf{P}_t$, the marginals $\mathbf{P}_t$ are uniquely determined for all $t \in [0, 1]$.

Statistical Estimator

Define the following fit function:

$$\mathrm{Fit}^{\lambda,\sigma}(g_\sharp \mathbf{R}_{t_1^T}, \ldots, g_\sharp \mathbf{R}_{t_T^T}) := \frac{1}{\lambda} \sum_{i=1}^{T} \Delta t_i \mathrm{DF}^\sigma(g_\sharp \mathbf{R}_{t_i^T}, \hat{\rho}_i^T)$$
$$\mathrm{DF}^\sigma(g_\sharp \mathbf{R}_{t_i^T}, \hat{\rho}_i^T) := H(\hat{\rho}_i^T | g_\sharp \mathbf{R}_{t_i^T} * \mathcal{N}_\sigma) + H(\hat{\rho}_i^T) + C,$$

where $H(\cdot|\cdot)$ is the Kullback-Leibler, $H(\cdot)$ is negative entropy, and C is a constant.

Our estimator is:

$$\mathcal{F}(\mathbf{R}) := \mathrm{Fit}^{\lambda,\sigma}(g_\sharp \mathbf{R}_{t_1^T}, \ldots, g_\sharp \mathbf{R}_{t_T^T}) + \tau H(\mathbf{R} | \mathbf{W}^{\Xi,\tau}), \tag{2}$$

where $\mathbf{W}^{\Xi,\tau}$ is the divergence-free path measure.

Theoretical Results

Theorem (consistency). *Suppose* $\mathbf{P}$ *is the SDE, Eq.* (1), *with initial condition* $\mathbf{P}_0 \in \mathcal{P}(\mathcal{X})$ *such that* $H(\mathbf{P}_0|\mathrm{vol}) < +\infty$. *Let* $\mathbf{R}^{T,\lambda,\sigma} \in \mathcal{P}(\Omega)$ *be the unique minimizer of* (2). *If* $\{t_i^T\}_{t \in [T]}$ *becomes dense in* $[0, 1]$, *then* $\lim_{\sigma \to 0, \lambda \to 0} \left(\lim_{T \to \infty} \mathbf{R}^{T,\lambda,\sigma}\right) = \mathbf{P}$ *almost surely.*

$\mathcal{F}$ is an optimization problem over the infinite-dimensional path space. We can introduce the following reduced objective on particle space:

$$F(\boldsymbol{\mu}) := \mathrm{Fit}^{\lambda,\sigma}(g_\sharp \boldsymbol{\mu}) + \sum_{i=1}^{T-1} \frac{1}{\Delta t_i} T_{\tau_i,\Xi}(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}^{(i+1)}) + \tau H(\boldsymbol{\mu}), \tag{3}$$

where $H(\boldsymbol{\mu}) = \sum_{i=1}^T \int \log(\boldsymbol{\mu}^{(i)}) d\boldsymbol{\mu}^{(i)}$ is the negative differential entropy of the family of measures $\boldsymbol{\mu}$ and

$$T_{\tau_i,\Xi}(\mu, \nu) := \min_{\gamma \in \Pi(\mu,\nu)} \int c_{\tau_i}^{\Xi}(x, y)\, d\gamma(x, y) + \tau_i H(\gamma | \mu \otimes \nu)$$

are entropic optimal transport plans with cost function $c_{\tau_i}^{\Xi}(x, y) := -\tau_i \log(p_{\tau_i}^{\Xi}(x, y))$ and p_t^{Ξ} is the transition probability density of $\mathbf{W}^{\Xi,\tau}$ over the time interval $[0, t]$.

We can show that minimizing Eq. (3) is equivalent to minimizing Eq. (2):

Theorem (representer). The following hold.

(i) *If* $\mathcal{F}$ *admits a minimizer* $\mathbf{R}^*$ *then* $(\mathbf{R}^*_{t_1^T}, \ldots, \mathbf{R}^*_{t_T^T})$ *is a minimizer for* F.

(ii) *If* F *admits a minimizer* $\boldsymbol{\mu}^* \in \mathcal{P}(\mathcal{X})^T$, *then a minimizer* $\mathbf{R}^*$ *for* $\mathcal{F}$ *is built as*

$$\mathbf{R}^*(\cdot) = \int_{\mathcal{X}^T} \mathbf{W}^{\Xi,\tau}(\cdot | x_1, \ldots, x_T)\, d\mathbf{R}_{t_1^T,\ldots,t_T^T}(x_1, \ldots, x_T),$$

where $\mathbf{W}^{\Xi,\tau}(\cdot|x_1, \ldots, x_T)$ *is the law of* $\mathbf{W}^{\Xi,\tau}$ *conditioned on passing through* $x_1, \ldots, x_T$ *at times* $t_1^T, \ldots, t_T^T$, *respectively and* $\mathbf{R}_{t_1^T,\ldots,t_T^T}$ *is the composition of the entropic optimal transport plans* γ_i *that minimize* $T_{\tau_i,\Xi}(\boldsymbol{\mu}^{*(i)}, \boldsymbol{\mu}^{*(i+1)})$, *for* $i \in [T-1]$.

Algorithm

We optimize F, Eq. (3), using mean-field Langevin (MFL) dynamics, which has exponential convergence guarantees [3]:

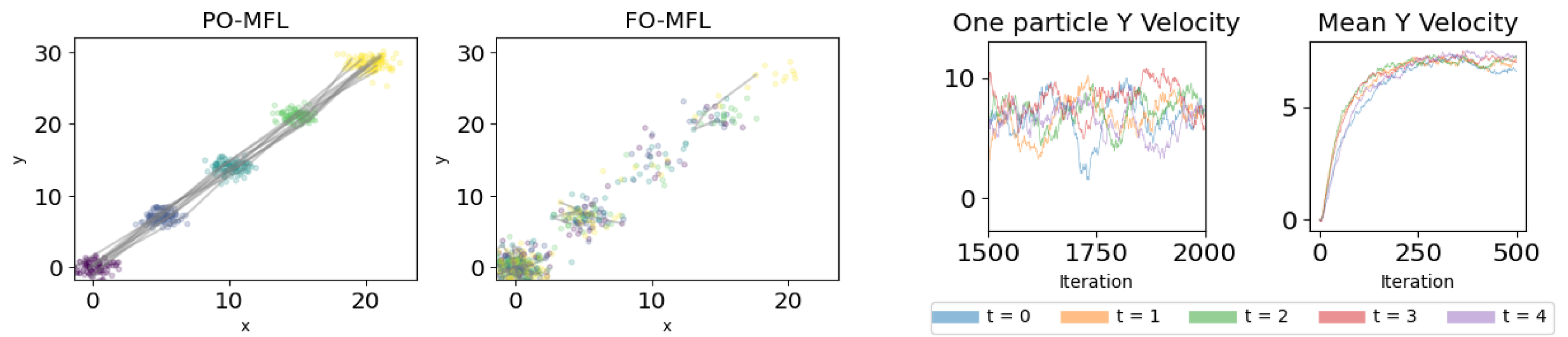
PO-MFL: framework for latent trajectory inference

Require: Collection of observations $(\hat{\rho}_1, \ldots, \hat{\rho}_T)$, collection of T time samples $(t_1^T, \ldots, t_T^T)$, velocity dynamics Ξ, number of iterations for MFL dynamics N, number of particles m, entropic OT parameter τ

Initialize m particles for each time: $(\hat{m}_1, \ldots, \hat{m}_T) \in \mathcal{X}^{m \times T}$
for N iterations **do**
  for $i \in [T-1]$ **do**
    $\Delta t_i := t_{i+1}^T - t_i^T$
    $C_i := \{C_{j,k}\}_{j,k=1}^m \leftarrow \frac{1}{2}\|\hat{m}_{i+1,k} - \hat{m}_{i,j} + \Delta t_i \Xi(t_i^T, \hat{m}_{i,j})\|^2$
    $\gamma_i \leftarrow \mathrm{Sinkhorn}(\hat{m}_i, \hat{m}_{i+1}, C_i, \tau \cdot \Delta t_i)$
  end for
  $\hat{\mathbf{m}} \leftarrow \mathrm{MFL}(\hat{\mathbf{m}}, \boldsymbol{\gamma}, \hat{\boldsymbol{\rho}})$ ▷ $\hat{\mathbf{m}} := (\hat{m}_1, \ldots, \hat{m}_T)$, etc.
end for
Output collection of particles $\hat{\mathbf{m}}$, trajectories $\gamma_{T-1} \circ \cdots \circ \gamma_1$

Synthetic Experiments

Synthetic experimens for a simple constant velocity model:



(left) PO-MFL (ours) recovers positions while FO-MFL (baseline) does not converge because velocity is too high to catch up. (right) PO-MFL recovers hidden velocity and verification of exponential convergence.

Applications (Future Work)

- Single-cell genomic data analysis: learning the distribution of cellular gene expression trajectories.
- Learning subject trajectory distributions from independent surveys at various times without the need to maintain a consistent panel of respondents.
- Private synthetic trajectory generation.

References

[1] H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger. Toward a mathematical theory of trajectory inference. *The Annals of Applied Probability*, 2024

[2] L. Chizat, S. Zhang, M. Heitz, and G. Schiebinger. Trajectory inference via mean-field Langevin in path space. *Advances in Neural Information Processing Systems*, 2022.

[3] L. Chizat. Mean-field Langevin dynamics: Exponential convergence and annealing. *Transactions on Machine Learning Research*, 2022.