Partially Observed Trajectory Inference using Optimal Transport and a Dynamics Prior

Anming Gu (Boston University)
Kristjan Greenewald (IBM Research)
Edward Chien (Boston University)

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 and provided a mean-field Langevin algorithm. We extend the guarantees to observable state-space models.

Problem Setup

Unobserved state vector in latent space $\mathcal{X}$ follows the SDE

$$dX_t = -\Xi(t, X_t)dt - \nabla\Psi(t, X_t)dt + \sqrt{\tau}dB_t$$

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

The state vector evolves over time $t \in [0,1]$ with initial distribution $\mathbf{P}_0$, which yields path law $\mathbf{P}$.

Observation space $\mathcal{Y}$: function $g : \mathcal{X} \to \mathcal{Y}$ is the observation function

T observation times with $0 \le t_1^T < \cdots < t_T^T \le 1$ and N_i^T i.i.d. samples:

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

Empirical distributions:

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

Problem: Given $(\hat{\rho}_1^T, \dots, \hat{\rho}_T^T)$, recover $\mathbf{P}$

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

Fit function:

$$\mathrm{Fit}^{\lambda,\sigma}(g_\sharp \mathbf{R}_{t_1^T}, \dots, g_\sharp \mathbf{R}_{t_T^T}) := \frac{1}{\lambda}\sum_{i=1}^{T} \Delta t_i \mathrm{DF}^\sigma(g_\sharp \mathbf{R}_{t_i^T}, \hat{\rho}_i^{T,h}),$$

$$\mathrm{DF}^\sigma(g_\sharp \mathbf{R}_{t_i^T}, \hat{\rho}_i^{T,h}) = H(\hat{\rho}_i^{T,h} | g_\sharp \mathbf{R}_{t_i^T} * \mathcal{N}_\sigma) + H(\hat{\rho}_i^{T,h}) + C,$$

where $H(\cdot|\cdot)$ is KL divergence and $H(\cdot)$ is negative entropy

Theoretical Results

Min-entropy estimator:

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

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

Main theorem:

Suppose $\mathbf{P}$ follows the SDE with initial condition $\mathbf{P}_0 \in \mathcal{P}(\mathcal{X})$ s.t. $H(\mathbf{P}_0|\mathrm{vol}) < +\infty$. Let $\mathbf{R}^{T,\lambda,h} \in \mathcal{P}(\Omega)$ be the unique minimizer of

$$\mathbf{R}^{T,\lambda,h} := \arg\min_{\mathbf{R} \in \mathcal{P}(\Omega)} \mathcal{F}(\mathbf{R}).$$

Then, it holds

$$\lim_{h \to 0, \lambda \to 0} \left(\lim_{T \to \infty} \mathbf{R}^{T,\lambda,h} \right) = \mathbf{P}.$$

The optimization on path-space is equivalent to one on particle-space:

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

where $T_{\tau_i,\Xi}$ are entropic OT costs, due to:

The following holds.

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

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

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

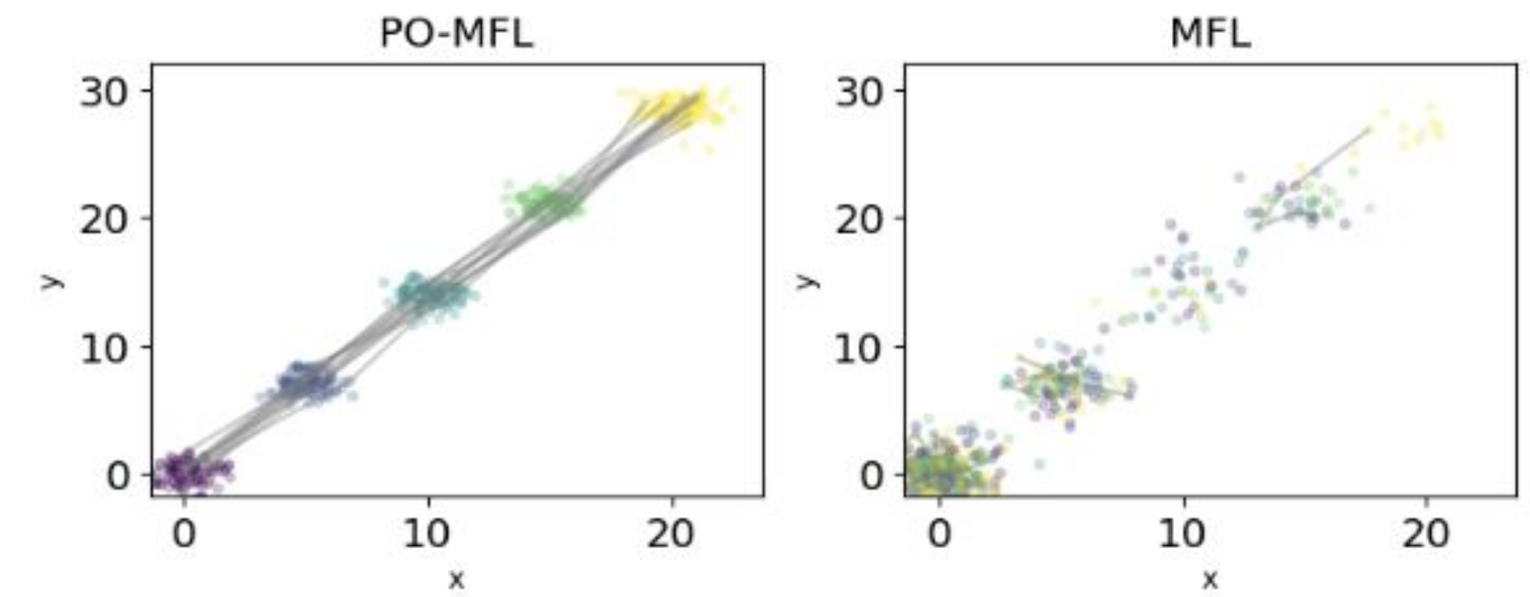
F can be optimized via mean-field Langevin dynamics [3].

Algorithm 1 PO-MFL: framework for latent trajectory inference

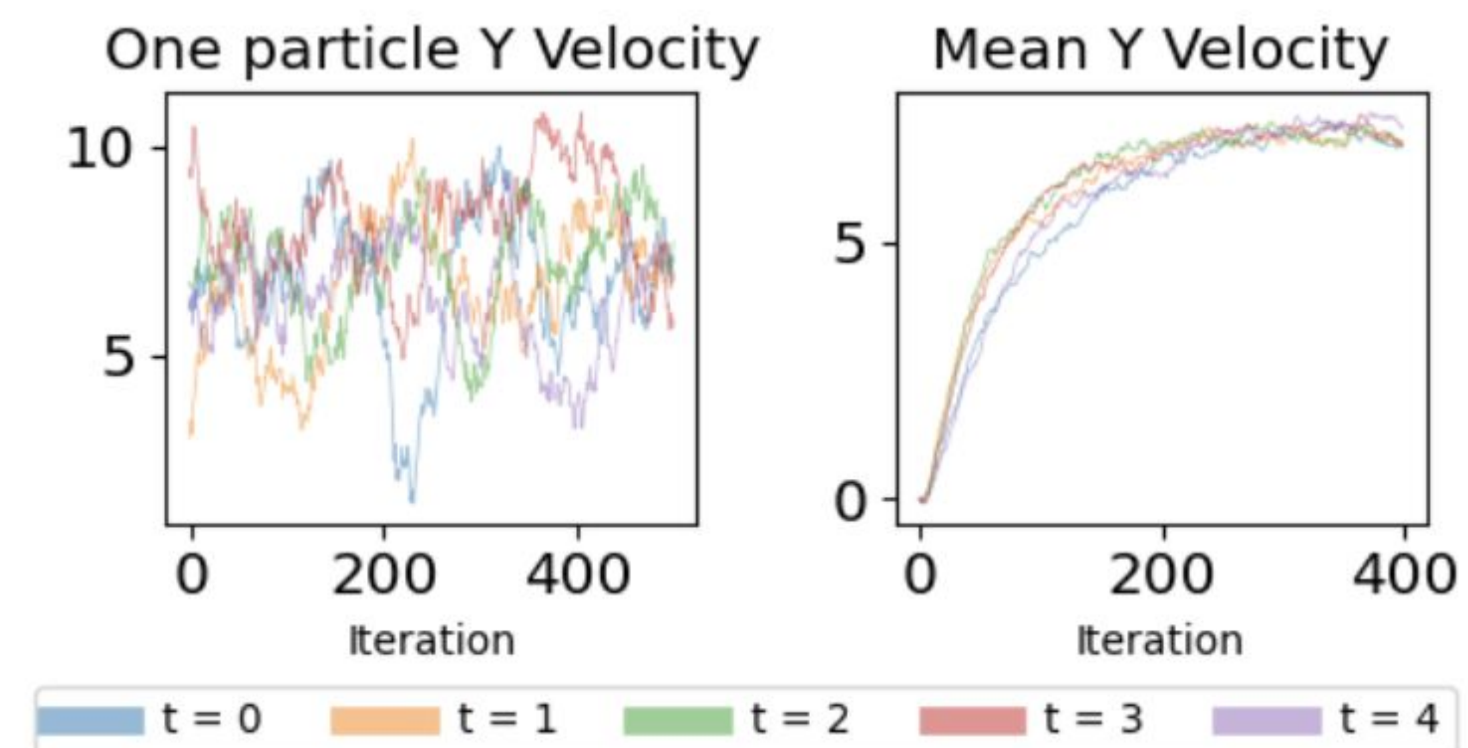
Require: Collection of observations $(\hat{\rho}_1, \dots, \hat{\rho}_t)$, collection of T time samples $(t_1^T, \dots, t_T^T)$, velocity dynamics Ξ, number of iterations for MFL dynamics N, number of particles m, entropic OT parameter λ

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

Synthetic Data Results



A simple constant velocity model. (left) PO-MFL algorithm that takes into account the velocity dynamics. (right) Baseline.



(left) Velocity of one particle at end of optimization. (right) Population velocity at start of optimization, showing 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] Lavenant et al (2024). *Towards a mathematical theory of trajectory inference*. The Annals of Applied Probability.

[2] Chizat et al (2022). *Trajectory inference via mean-field Langevin in path space*. Neural Information Processing Systems.

[3] Chizat (2022). *Mean-field Langevin dynamics: exponential convergence and annealing*. Transactions on Machine Learning Research.