Latent Trajectory Inference with Drift Prior via Optimal Transport

Anming Gu

May 3rd, 2024

Table of Contents

Motivation: computational biology applications

Goal: understand biological processes

Issue: we cannot observe full cell development process

Data consists of population snpashots at different time points

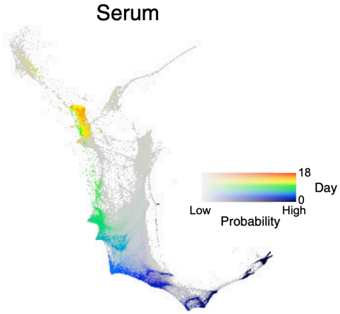


Figure from Schiebinger et al., 2019

What is trajectory inference?

Let $\mathcal{X}$ be the ambient space and $\Omega = C([0,1] : \mathcal{X})$ be the path space

Goal: estimate the ground truth stochastic process $\mathbf{P} \in \mathcal{P}(\Omega)$

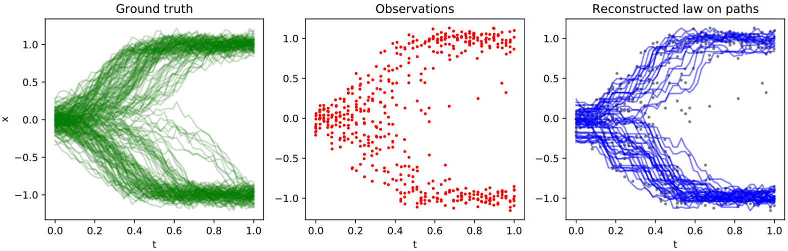


Figure from Lavenant et al., 2021

Mathematical model of trajectory inference

Let $X_t \in \mathcal{X}$ be an unobserved state vector evolving according to the following SDE for $t \in [0, 1]$:

$$dX_t = -\Xi(t, X_t)dt - \nabla\Psi(t, X_t)dt + \sqrt{\tau}dB_t \tag{1}$$

- initial condition $X_0 \sim \mu_0$
- divergence-free velocity prior $\Xi \in C([0, 1] \times \mathcal{X} : \mathcal{X})$ is *known*
- potential $\Psi \in C^2([0, 1] \times \mathcal{X} : \mathbb{R})$ is *unknown*
- $\tau > 0$ is the variance, $\{B_t\}$ is a standard Brownian motion

This is our ground truth $\mathbf{P} \in \mathcal{P}(\Omega)$

Measurement Model

Smooth function $g : \mathcal{X} \to \mathcal{Y}$ transforming X_t into the observation space $\mathcal{Y}$:

$$Y_t = g(X_t)$$

T observation times with $0 \leq t_1^T < \cdots < t_T^T \leq 1$, and we observe N_i^T i.i.d. samples from the marginal distribution of Y_{t_i}:

$$\{Y_{i,j}^T\}_{j=1}^{N_i^T} \overset{\text{i.i.d.}}{\sim} g_\sharp \mathbf{P}_{t_i^T} := \mathbf{Q}_{t_i^T}.$$

Smooth empirical distribution by h-wide heat kernel Φ_h:

$$\hat{\rho}_i^T = \Phi_h \left(\frac{1}{N_i^T} \sum_{i=1}^{N_i^T} \delta_{Y_{i,j}^T} \right)$$

Goal: recover $\mathbf{P}$ from $(\hat{\rho}_1^T, \ldots, \hat{\rho}_T^T)$ and the known velocity field Ξ

Observability assumption

Ψ is unknown, but restricted to a class $\mathcal{C}_\Psi$.

$(g, \Xi, \mathcal{C}_\Psi)$ is $\mathcal{C}_\Psi$-*marginal-observable* if, given g, Ξ, σ, and all marginals $\mathbf{Q}_t = g_\sharp \mathbf{P}_t$ of Y_t for all $t \in [0,1]$, the marginals $\mathbf{P}_t$ of X_t are uniquely determined for all $t \in [0,1]$

With this assumption, we can infer the latent dynamics solely from the marginals $\mathbf{Q}_t$

Why is our setting important?

Goal: recover **P** from $(\hat{\rho}_1^T, \ldots, \hat{\rho}_T^T)$ and the known velocity field Ξ

Applications:

- More robust optimization using velocity prior
- Smoother trajectories and more accurate prediction of final particle location
- Privacy: don't need to release full data
- May be useful for studying diffusion models

Table of Contents

Data-fitting term

Let $\Delta t_i := t_{i+1}^T - t_i^T$. Fit function: $\mathrm{Fit}^{\lambda,\sigma} : \mathcal{P}(\mathcal{Y})^T \to \mathbb{R}$:

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

$$\begin{aligned}\mathrm{DF}^\sigma(g_\sharp \mathbf{R}_{t_i^T}, \hat{\rho}_i^{T,h}) &:= \int_{\mathcal{Y}} -\log\left[\int_{\mathcal{X}} \exp\left(-\frac{\|g(x)-y\|^2}{2\sigma^2}\right) d\mathbf{R}_{t_i^T}(x)\right] d\hat{\rho}_i^{T,h}(y) \\ &= H(\hat{\rho}_i^{T,h} | g_\sharp \mathbf{R}_{t_i^T} * \mathcal{N}_\sigma) + H(\hat{\rho}_{t_i^T}) + C,\end{aligned}$$

where $\mathcal{N}_\sigma$ is the Gaussian kernel with variance σ^2, $C > 0$ is a constant

- Negative log-likelihood under the noisy observation model $\hat{Y}_{i,j}^T = g(X_{i,j}^T) + \sigma Z_{i,j}$, where $\hat{Y}_{i,j}^T$ is the observation and $Z_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0, I)$.
- DF^σ is jointly convex in $(\mathbf{R}_{t_i^T}, \hat{\rho}_i^{T,h})$ and linear in $\hat{\rho}_i^{T,h}$.

Chizat et al., 2022

Min-entropy estimator

Functional $\mathcal{F} : \mathcal{P}(\Omega) \to \mathbb{R}$

$$\mathcal{F}(\mathbf{R}) := \mathrm{Fit}^{\lambda,\sigma}(\mathbf{Q}_{t_1^T}, \ldots, \mathbf{Q}_{t_T^T}) + \tau H(\mathbf{R}|\mathbf{W}^{\Xi,\tau}), \quad \mathbf{R}^{T,\lambda,h} := \arg\min \mathcal{F}(\mathbf{R})$$

- $\mathbf{W}^{\Xi,\tau} \in \mathcal{P}(\Omega)$ is the law of the SDE $dZ_t = -\Xi(t, Z_t)\,dt + \sqrt{\tau}\,dB_t$
- $H(\mu|\nu) = \int \log(d\mu/d\nu)\,d\mu$ is relative entropy
- Fit term on previous slide

Theorem (Consistency, Lavenant et al., 2021)

If $\{t_i^T\}_{i\in[T]}$ *becomes dense in* $[0,1]$ *as* $T \to \infty$*,*

$$\lim_{\lambda,h\to 0} \lim_{T\to\infty} \mathbf{R}^{T,\lambda,h} = \mathbf{P}$$

weakly, a.s.

High level ideas for proof of consistency

Recall $\mathbf{P}$ follows the SDE $dX_t = -\Xi(t, X_t)dt - \nabla\Psi(t, X_t)dt + \sqrt{\tau}dB_t$ and $\mathbf{W}^{\Xi,\tau}$ follows the SDE $dZ_t = -\Xi(t, Z_t) + \sqrt{\tau}dB_t$

- Drift term in Z_t cancels out that of X_t in the analysis, e.g. can check via Girsanov
- This yields analogous statements as in Lavenant et al., 2021

Main results follow using Γ-convergence theory and analysis:

-

Table of Contents

Our entropic optimal transport problem

Goal: reduce the problem over the space $\mathcal{P}(\mathcal{X}^T)$ to use the mean-field Langevin (MFL) dynamics

Let $\tau_i := \Delta t_i \cdot \tau$ and consider the entropic OT problem:

$$\begin{aligned} T_{\tau_i,\Xi}(\mu,\nu) &:= \min_{\gamma\in\Pi(\mu,\nu)} \int c^{\Xi}_{\tau_i}(x,y)\, d\gamma(x,y) + \tau_i H(\gamma|\mu\otimes\nu) \\ &= \min_{\gamma\in\Pi(\mu,\nu)} \tau_i H(\gamma|p^{\Xi}_{\tau_i}\mu\otimes\nu) \end{aligned}$$

- set of transport plans $\Pi(\mu,\nu)$, e.g. probability measures in $\mathcal{P}(\mathcal{X}\times\mathcal{X})$ with marginals μ,ν
- cost function $c^{\Xi}_{\tau_i}(x,y) := -\Delta t_i \log(p^{\Xi}_{\tau_i}(x,y))$
- $p^{\Xi}_{\tau_i}$ transition probability density of $\mathbf{W}^{\Xi}$

Chizat et al., 2022

Representer theorem

Optimization over $\mathcal{P}(\Omega)$:

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

Reduced optimization over $\mathcal{P}(\mathcal{X})^T$:

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

Theorem (Chizat et al., 2022)

A minimizer for $\mathcal{F}$ can be built from a minimizer for F.

Composition of optimal transport plans:

$$\mathbf{R}_{t_i,\ldots,t_T}(dx_1, \ldots, dx_T) = \gamma_{1,2}(dx_1, dx_2)\gamma_{2,3}(dx_3|x_2)\cdots\gamma_{T-1,T}(dx_T|x_{T-1})$$

Approximation of the entropic OT problem

We still cannot solve F! Why?

$p^{\Xi}_{\tau_i}$ is generally not well-defined

Idea: approximate $T_{\tau_i,\Xi}(\mu,\nu)$ using an Euler-Maruyama discretization

Consider two time points $t_1 \leq t_2$. Define

$$\Xi^{\Delta t}_{\sharp}\mu_{t_1} := \mu_{t_1} - \Xi(t_1, \mu_{t_1}) \cdot \Delta t$$

Consider $T_{\tau_i}(\Xi^{\Delta t}_{\sharp}\mu_{t_1}, \mu_{t_2})$ instead of $T_{\tau_i,\Xi}(\mu_{t_1}, \mu_{t_2})$, where

$$T_{\tau_i}(\mu,\nu) := \min_{\gamma \in \Pi(\mu,\nu)} \tau_i H(\gamma | p_{\tau_i}\mu \otimes \nu)$$

is the entropic OT cost corresponding to the transition probability density of the Brownian motion $p_t(x,y)$

Approximation of the entropic OT problem (cont.)

Theoretical justification:

Proposition

Assume $\mathcal{X}$ is a bounded domain, e.g. $\operatorname{diam}\mathcal{X} < +\infty$. *Let* $\Delta t := t_2 - t_1$ *and* $\tau_i := \tau \Delta t$. *Define* $\Xi_\sharp^{\Delta t} x := x - \Xi(t_1, x) \cdot \Delta t$. *We have*

$$\lim_{\Delta t \to 0} \int_{\mathcal{X} \times \mathcal{X}} \left| \log(p_{\tau_i}^{\Xi}(x, y)) - \log(p_{\tau_i}(\Xi_\sharp^{\Delta t} x, y)) \right| dx\, dy = 0.$$

Proof idea: use triangle inequality, Taylor approximation, dominated convergence, and fact that transition kernel is Dirac delta in the limit.

No rate of convergence

Discussion of approximation

- Computationally, consider: $T_{\tau_i}(\Xi_\sharp^{\Delta t/2}\mu_{t_1}, \Xi_\sharp^{-\Delta t/2}\mu_{t_2})$.
- Varadhan's approximation:

$$\tilde{c}_{\tau_i}^{\Xi}(x,y) \approx \frac{1}{2}\left\| y + \frac{\Delta t}{2}\Xi(t_2,y) - x + \frac{\Delta t}{2}\Xi(t_1,x)\right\|^2,$$

  which holds for τ_i small
- Consistency result: justifies using Ξ in entropic OT problem
- Intuition for robustness: $\mathbb{E}[|\Xi_\sharp^{\Delta t/2}\mu_{t_1} - \Xi_\sharp^{-\Delta t/2}\mu_{t_2})|] \approx 0$ even if the particles move a large distance

Table of Contents

Mean-field Langevin dynamics

For convex $G : \mathcal{P}(\mathcal{X}) \to \mathbb{R}_{\geq 0}$, MFL dynamics solves the following optimization problem:

$$\min_{\mu \in \mathcal{P}_2(\mathcal{X})} F_\tau(\mu) := G(\mu) + H(\mu)$$

Let $V[\mu] := \frac{\delta G}{\delta \mu}(\mu) \in C^1(\mathcal{X})$ be the *first variation* of G:

$$\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}[G((1-\epsilon)\mu + \epsilon)) - G(\mu)] = \int_{\mathcal{X}} V[\mu](x)\, d(\nu - \mu)(x)$$

for all μ, ν.

Chizat et al., 2022

Noisy particle gradient descent

Optimization by running noisy particle gradient descent on $G_m : (\mathcal{X}^m)^T \to \mathbb{R}$ defined as $G_m(\hat{X}) := G(\hat{\boldsymbol{\mu}}_{\hat{X}})$, where

$$\hat{\boldsymbol{\mu}}^{(i)}_{\hat{X}} = \frac{1}{m} \sum_{j=1}^{m} \delta_{\hat{X}^{(i)}_j}.$$

Optimization procedure is:

$$\begin{cases} \hat{X}^{(i)}_j[k+1] = \hat{X}^{(i)}_j[k] - \eta \nabla V^{(i)}[\hat{\boldsymbol{\mu}}[k]](\hat{X}^{(i)}_j[k]) + \sqrt{2\eta(\tau+\epsilon)} Z^{(i)}_{j,k}, \\ \hat{\boldsymbol{\mu}}^{(i)}[k] = \frac{1}{m} \sum_{j=1}^{m} \delta_{\hat{X}^{(i)}_j[k]} \quad i \in [T], \end{cases}$$

- $\hat{X}^{(i)}_j[0] \overset{i.i.d.}{\sim} \boldsymbol{\mu}^{(i)}_0$
- $\eta > 0$ is a step-size
- $Z^{(i)}_{j,k}$ are i.i.d. standard Gaussian variables

Chizat et al., 2022

Exponential convergence

Theorem (Chizat, 2022)

Let $\boldsymbol{\mu}_0 \in \mathcal{P}(\mathcal{X})^T$ be such that $F(\boldsymbol{\mu}_0) < \infty$. Then for $\epsilon \geq 0$, there exists a unique solution $(\boldsymbol{\mu}_s)_{s \geq 0}$ to the MFL dynamics. For $\epsilon > 0$, $\mathcal{X}$ the d-torus, and moreover assuming that $\boldsymbol{\mu}_0$ has a bounded absolute log-density, it holds

$$F_\epsilon(\boldsymbol{\mu}_s) - \min F_\epsilon \leq e^{-Cs}(F_\epsilon(\boldsymbol{\mu}_0) - \min F_\epsilon),$$

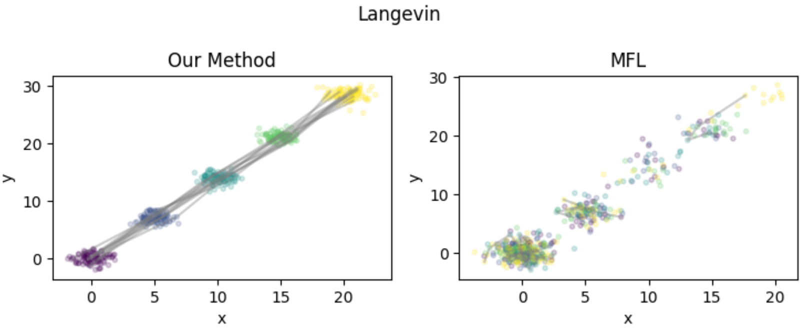
where $C = \beta e^{-\alpha/\epsilon}$ for some $\alpha, \beta > 0$ independently of $\boldsymbol{\mu}$ and ϵ.

Taking ϵ_s decaying small enough, we have $\boldsymbol{\mu}_s$ converges weakly to the min-entropy estimator $\boldsymbol{\mu}^*$.

Chizat et al., 2022; Chizat, 2022

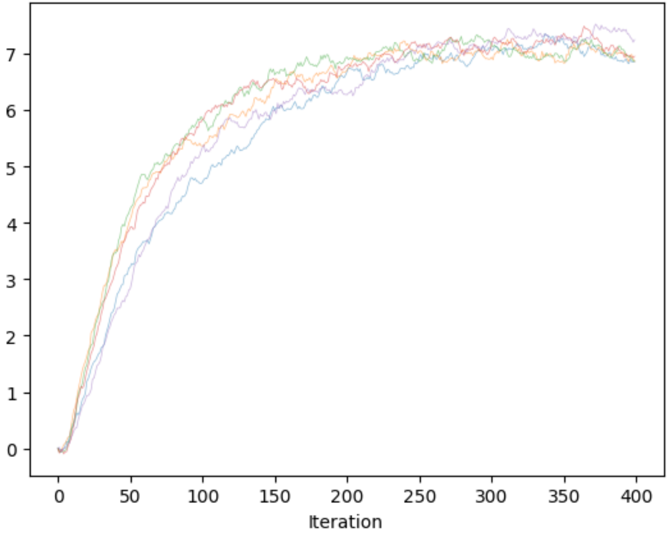
Table of Contents

A simple velocity model: robustness

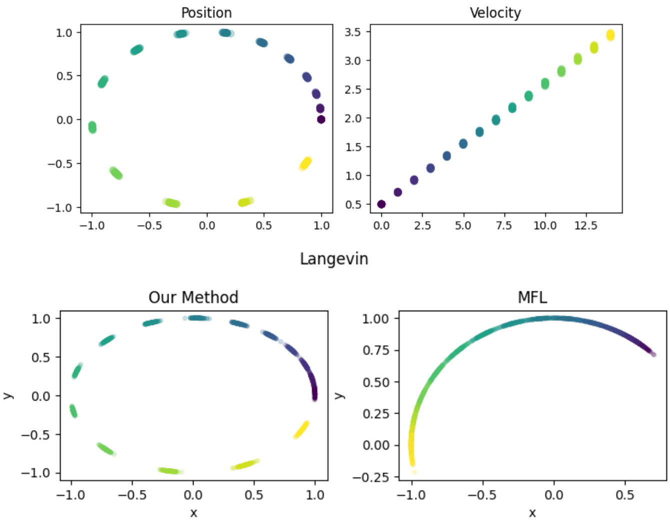


- Initial condition is at the origin
- x velocity: 5, y velocity: 7
- MFL fails to converge

A simple velocity model: exponential convergence



A circular motion model: recovered position



A circular motion model: recovered velocity

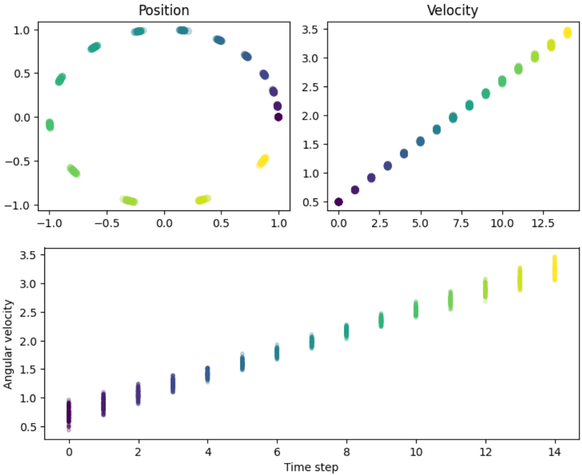


Table of Contents

Future work

Conjecture

For every $t_1 < t_2$, *with* $\Delta t := t_2 - t_1$ *sufficiently small, we have*

$$|T_{\tau_i,\Xi}(\mu_{t_1}, \mu_{t_2}) - T_{\tau_i}(\Xi_\sharp^{\Delta t}\mu_{t_1}, \mu_{t_2})| = O(\Delta t)$$

and

$$|H(\gamma_\Xi|\mathbf{W}_{t_1,t_2}^{\Xi,\tau}) - H(\gamma|\mathbf{W}_{t_1,t_2}^{\tau})| = O(\Delta t),$$

where γ_Ξ *and* γ *are the corresponding optimal transport plan to* $T_{\tau_i,\Xi}(\mu_{t_1}, \mu_{t_2})$ *and* $T_{\tau_i}(\Xi_\sharp^{\Delta t}\mu_{t_1}, \mu_{t_2})$, *respectively.*

- Statistical properties of the estimator.
- Relaxed assumptions on g, Ξ
- Empirical validation of predicting outcomes of individuals

References

- Schiebinger, Shu, Tabaka, Cleary, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming, 2019.
- Lavenant, Zhang, Kim, Schiebinger, 2021. Towards a mathematical theory of trajectory inference.
- Chizat, Zhang, Heitz, Schiebinger, 2022. Trajectory inference vi mean-field Langevin in path space.
- Chizat, 2022. Mean Field Langevin Dynamics: Exponential convergence and annealing

Conclusion