Trajectory Inference via Mean-Field Langevin in Path Space

Lénaïc Chizat*
Based on joint works with: Guillaume Carlier, Matthieu Heitz, Maxime Laborde, Geoffrey Schiebinger, Stephen Zhang

IBM research 2022 - New-York

*Institute of Mathematics, EPFL, DOLA chair

Outline

Trajectory Inference & Min-entropy Estimator

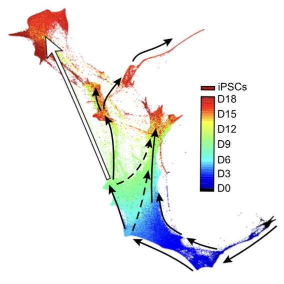
Motivation: process single-cell RNA sequencing data

Data

Gene expression level of individual cells sampled at several times

Goal

Understand biological processes (development, reprogramming) : genealogy of cells, role of genes, effect of interventions, etc.

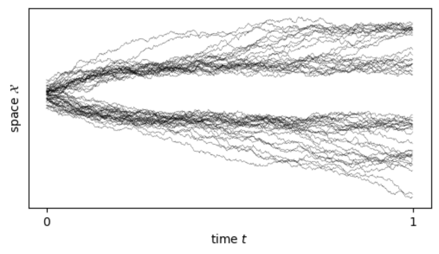


Taken from (Schiebinger et al. 2019)

See also (Tong et al. 2020), (Farrell et al. 2018),...

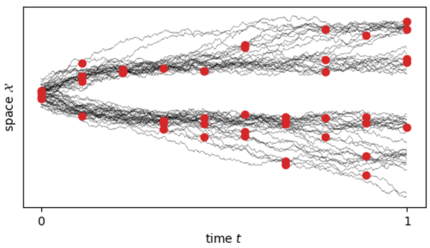
Mathematical Model of Trajectory Inference (I)

- Ambient space $\mathcal{X} \subset \mathbb{R}^d$ convex compact
- Path-space $\Omega := \mathcal{C}([0,1];\mathcal{X})$
- **Goal:** estimate the population dynamics $P \in \mathcal{P}(\Omega)$



Mathematical Model of Trajectory Inference (I)

- Ambient space $\mathcal{X} \subset \mathbb{R}^d$ convex compact
- Path-space $\Omega := \mathcal{C}([0,1];\mathcal{X})$
- **Goal:** estimate the population dynamics $P \in \mathcal{P}(\Omega)$



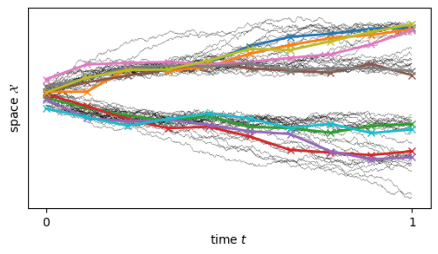
Measurement Model

Observe $(X_{t_i,j})_{i\in[T],j\in[n_i]}$ for $0 \leq t_1,\dots,t_T \leq 1$ and $n_i \geq 1$.

- independent realizations for all couples (i,j)
- Snapshots: $\hat{\mu}_{t_i} := \frac{1}{n_i}\sum_{j=1}^{n_i} \delta_{X_{t_i,j}}$

Mathematical Model of Trajectory Inference (I)

- Ambient space $\mathcal{X} \subset \mathbb{R}^d$ convex compact
- Path-space $\Omega := \mathcal{C}([0,1];\mathcal{X})$
- **Goal:** estimate the population dynamics $P \in \mathcal{P}(\Omega)$



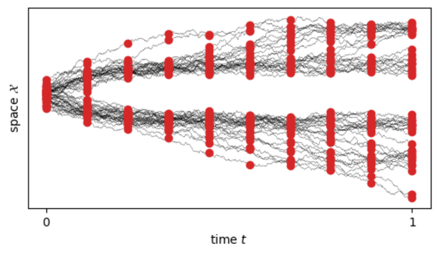
Measurement Model

Observe $(X_{t_i,j})_{i\in[T],j\in[n_i]}$ for $0 \le t_1, \ldots, t_T \le 1$ and $n_i \ge 1$.

- independent realizations for all couples (i,j)
- Snapshots: $\hat{\mu}_{t_i} := \frac{1}{n_i}\sum_{j=1}^{n_i} \delta_{X_{t_i,j}}$

Mathematical Model of Trajectory Inference (I)

- Ambient space $\mathcal{X} \subset \mathbb{R}^d$ convex compact
- Path-space $\Omega := \mathcal{C}([0,1];\mathcal{X})$
- **Goal:** estimate the population dynamics $P \in \mathcal{P}(\Omega)$



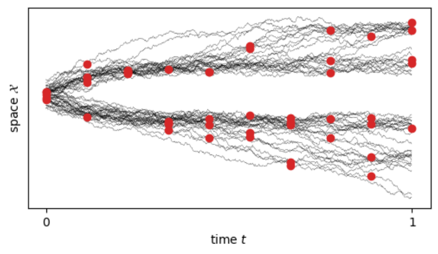
Measurement Model

Observe $(X_{t_i,j})_{i\in[T],j\in[n_i]}$ for $0 \leq t_1,\ldots,t_T \leq 1$ and $n_i \geq 1$.

- independent realizations for all couples (i,j)
- Snapshots: $\hat{\mu}_{t_i} := \frac{1}{n_i}\sum_{j=1}^{n_i} \delta_{X_{t_i,j}}$

Mathematical Model of Trajectory Inference (I)

- Ambient space $\mathcal{X} \subset \mathbb{R}^d$ convex compact
- Path-space $\Omega := \mathcal{C}([0,1];\mathcal{X})$
- **Goal:** estimate the population dynamics $P \in \mathcal{P}(\Omega)$



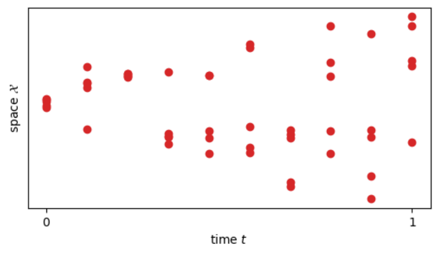
Measurement Model

Observe $(X_{t_i,j})_{i\in[T],j\in[n_i]}$ for $0 \leq t_1, \dots, t_T \leq 1$ and $n_i \geq 1$.

- independent realizations for all couples (i,j)
- Snapshots: $\hat{\mu}_{t_i} := \frac{1}{n_i}\sum_{j=1}^{n_i} \delta_{X_{t_i,j}}$

Mathematical Model of Trajectory Inference (I)

- Ambient space $\mathcal{X} \subset \mathbb{R}^d$ convex compact
- Path-space $\Omega := \mathcal{C}([0,1];\mathcal{X})$
- **Goal:** estimate the population dynamics $P \in \mathcal{P}(\Omega)$



Measurement Model

Observe $(X_{t_i,j})_{i\in[T],j\in[n_i]}$ for $0 \leq t_1,\ldots,t_T \leq 1$ and $n_i \geq 1$.

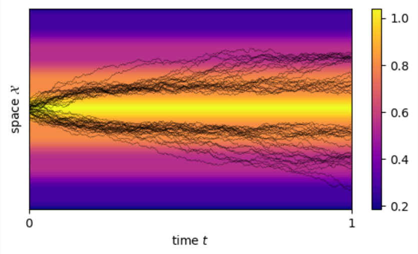
- independent realizations for all couples (i,j)
- Snapshots: $\hat{\mu}_{t_i} := \frac{1}{n_i}\sum_{j=1}^{n_i} \delta_{X_{t_i,j}}$

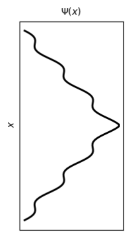
Mathematical Model of Trajectory Inference (II)

Potential-driven Ito Diffusion Model

$$dX_t = -\nabla\Psi(t, X_t)\,\mathrm{d}t + \sqrt{\tau}\,\mathrm{d}B_t, \quad X_0 \sim \mu_0$$

- potential $\Psi \in \mathcal{C}^2([0,1], \mathbb{R}^d)$ unknown
- temperature $\tau > 0$ known, B (reflected) Brownian motion
- characterizes $P \in \mathcal{P}(\Omega)$





$\rightsquigarrow$ some works focus on recovering Ψ, a different problem e.g. (Bunne et al. '21), (Hashimoto, '16), (Tong et al. '20)

Min-entropy estimator & Consistency

Estimator Min-entropy relative to Wiener measure

$$R^* := \operatorname*{argmin}_{R \in \mathcal{P}(\Omega)} \mathcal{F}(R), \quad \mathcal{F}(R) := \mathrm{Fit}_{\lambda,\sigma}(R_{t_1}, \ldots, R_{t_T}) + \tau H(R|W^\tau)$$

- $W^\tau \in \mathcal{P}(\Omega)$ is the law of the Brownian motion at temperature τ (reversible, reflected, on $\mathcal{X}$)
- $H(\mu|\nu) = \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\,\mathrm{d}\mu$ is the relative entropy
- see next slide for $\mathrm{Fit}_{\lambda,\sigma}$

Theorem [Lavenant et al. 2021]

If $(t_i)_{i \in [T]}$ becomes dense in $[0,1]$ as T grows, then

$$\lim_{\lambda,\sigma \to 0} \lim_{T \to \infty} R^* = P \quad \text{weakly, a.s.}$$

Lavenant, Zhang, Kim, Schiebinger (2021). *Towards a mathematical theory of trajectory inference.*

Data fitting term

$$\mathrm{Fit}_{\lambda,\sigma}(R_{t_1}, \ldots, R_{t_T}) := \frac{1}{\lambda} \sum_{i=1}^{T} (\Delta t_i) \widetilde{\mathrm{Fit}}_\sigma(R_{t_i} | \hat{\mu}_{t_i})$$

Log-likelihood fitting loss

Let $\widetilde{\mathrm{Fit}}_\sigma$ be the neg-log-likelihood under noisy observation model

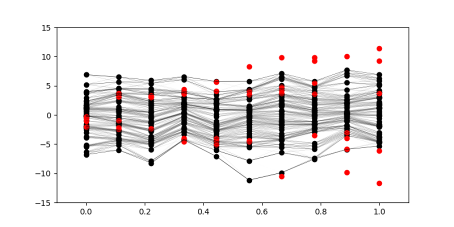
$$\hat{X}_{t_i,j} = X_{t_i,j} + \sigma Z_{i,j}, \quad X_{t_i,j} \sim R_{t_i}, \quad Z_{i,j} \sim \mathcal{N}(0,1)$$

$$\widetilde{\mathrm{Fit}}_\sigma(R_{t_i} | \hat{\mu}_{t_i}) := \int -\log \left(\int \exp\left(-\frac{\|x-y\|^2}{2\sigma^2} \right) \mathrm{d}R_{t_i}(x) \right) \mathrm{d}\hat{\mu}_{t_i}(y)$$

- Linear in $\hat{\mu}_{t_i}$
- Convex, smooth in R: as nice as one could hope

- Well-posed convex optimization problem over $\mathcal{P}(\Omega)$
- Discretize then optimize approach is tractable...
  $\rightsquigarrow$ reduction from $\mathcal{P}(\Omega)$ to $\mathcal{P}(\mathcal{X})^T$ thanks to the Markovian structure (Benamou et al. 2018), (Lavenant et al. 2021)
- ... but not satisfying (curse of dimensionality)

Can we design a free-support method that computes the min-entropy estimator R^*?



Chizat, Zhang, Heitz, Schiebinger (2022). *Trajectory Inference via Mean-field Langevin in Path Space.*

Reduced Formulation

Entropic Optimal Transport

Let $\Pi(\mu, \nu)$ be the set of transport plans between $\mu, \nu \in \mathcal{P}(\mathcal{X})$, i.e. probability measures in $\mathcal{P}(\mathcal{X} \times \mathcal{X})$ with marginals μ and ν.
$\mathcal{X} \subset \mathbb{R}^d$ compact.

Entropic Optimal Transport

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

where $c_\tau(x, y) \underset{\tau \to 0}{\longrightarrow} \frac{1}{2}\|y - x\|^2$ is the log-heat-kernel on $\mathcal{X}$.

- differentiable in (μ, ν)
- first variation given by the "stable" dual potentials (φ, ψ)
- ϵ-approximation in $O(n^2/\epsilon)$ using Sinkhorn's algorithm, if μ and ν have n atoms

A “representer theorem”

Path-space formulation over $\mathcal{P}(\Omega)$:

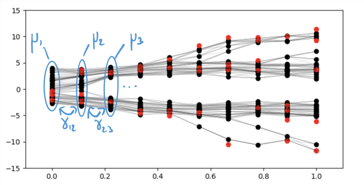
$$\mathcal{F}(R) := \mathrm{Fit}(R_{t_1}, \dots, R_{t_T}) + \tau H(R|W^\tau)$$

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

$$F(\boldsymbol{\mu}) := \underbrace{\mathrm{Fit}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T) + \sum_{i=1}^{T-1} \frac{1}{\Delta t_i} T_{\tau \Delta t_i}(\boldsymbol{\mu}_i, \boldsymbol{\mu}_{i+1})}_{G(\boldsymbol{\mu})} + \tau \underbrace{\sum_{i=1}^{T} H(\boldsymbol{\mu}_i)}_{H(\boldsymbol{\mu})}$$

Theorem

There is a computable bijection between minimizers of $\mathcal{F}$ and F.



$$R^*_{t_1,\dots,t_T}(dx_1, \dots, dx_T) = \boldsymbol{\mu}_1(dx_1)\gamma_{2|1}(dx_2) \dots \gamma_{T|T-1}(dx_T)$$

In (Chizat, Zhang, Heitz, Schiebinger, 2022)
Adapted from (Benamou et al. 2019), (Lavenant et al. 2020)

Mean-Field Langevin & Exponential Convergence

(Overdamped) Langevin Dynamics : quick primer (I)

- **Goal**: given $V \in \mathcal{C}^2(\mathcal{X})$, sample from $\propto e^{-V/\tau}$, $\tau > 0$.
- **Noisy GD**:

$$X_{k+1} = -\eta \nabla V(X_k) + \sqrt{2\tau\eta} Z_k, \quad X_0 \sim \mu_0, \quad Z_k \overset{iid}{\sim} \mathcal{N}(0, I)$$

- As $\eta \to 0$, converges in law to a **Langevin Dynamics** ($t = k\eta$):

$$dX_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2\tau} dB_t, \quad X_0 \sim \mu_0, \quad B_t \text{ Brownian process}$$

- Moreover $\mu_t = \mathrm{Law}(X_t)$ follows the **Fokker-Planck equation**:

$$\partial_t \mu_t = \underbrace{\nabla \cdot (\mu_t \nabla V)}_{\text{drift}} + \underbrace{\tau \Delta \mu_t}_{\text{diffusion}}, \quad \mu_0 \text{ given}$$

NB: do not confuse optimization time vs biological time

(Overdamped) Langevin Dynamics : quick primer (II)

Interpretation: Wasserstein gradient flow of

$$F_\tau(\mu) := \int V \, \mathrm{d}\mu + \tau H(\mu) = H(\mu|\mu^*_\tau)$$

where $\mu^*_\tau \propto e^{-V/\tau} \in \mathcal{P}(\mathcal{X})$, $H(\mu) = \int \log(\mathrm{d}\mu/\mathrm{d}x)\,\mathrm{d}\mu$ is the neg-entropy and $H(\mu|\nu) = \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\,\mathrm{d}\mu$ is the relative entropy

Theorem [Holley, Kusuoka, Stroock, 1989] and many more

- F_τ admits a unique minimizer $\mu^*_\tau \propto e^{-V/\tau}$.
- Assume that μ^*_τ satisfies a ρ_τ- *log-Sobolev inequality*, then

$$F_\tau(\mu_t) - F_\tau(\mu^*_\tau) \le e^{-2\tau\rho_\tau t}\big(F_\tau(\mu_0) - F_\tau(\mu^*_\tau)\big).$$

$\rightsquigarrow$ Let us generalize this result to a much larger class of dynamics with non-linear drift (interacting particles)

Setting for Mean-field Langevin

Let $G : \mathcal{P}(\mathcal{X}) \to \mathbb{R}_+$ a convex function (here $\mathcal{X} = \mathbb{R}^d$ allowed).

Optimization problem

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

Let $V[\mu] := \frac{\delta G}{\delta \mu}(\mu) \in \mathcal{C}^1(\mathcal{X})$ the first variation of G, i.e. $\forall \mu$,

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1}\big(G((1-\epsilon)\mu + \epsilon\nu) - G(\mu)\big) = \int_{\mathcal{X}} V[\mu](x)\,\mathrm{d}(\nu - \mu)(x), \quad \forall \nu$$

Assumptions

- assume that F_τ admits a minimizer μ^*_τ, let $F^*_\tau = F_\tau(\mu^*_\tau)$
- assume V exists and has a Lipschitz gradient, i.e. $\exists L > 0$ s.t.

$$\|\nabla V[\mu](x) - \nabla V[\mu'](x')\| \leq L\big(W_2(\mu, \mu') + \|x - x'\|\big), \;\; \forall \mu, \mu', x, x'$$

Particle gradient flow and its mean-field limit

Noisy Gradient Flow (Evolution of m particles)

$$dX_t^{(i)} = -\nabla V[\hat{\mu}_t](X_t^{(i)})\,\mathrm{d}t + \sqrt{2\tau}dB_{t,i} \quad \text{where} \quad \hat{\mu}_t = \frac{1}{m}\sum_{i=1}^{m}\delta_{X^{(i)}(t)}$$

Proposition (Mean-Field limit [Mc-Kean, Kac... in the 60s])

As $m \to +\infty$, the $(X_t^{(i)})$ converges in distribution to iid draws from

$$dX_t = -\nabla V[\mu_t](X_t)\,\mathrm{d}t + \sqrt{2\tau}dB_t \quad \text{where} \quad \mu_t = \mathrm{Law}(X_t)$$

Mean-Field Langevin Dynamics

$$\partial_t \mu_t = \nabla\cdot(\mu_t \nabla V[\mu_t]) + \tau\Delta\mu_t$$

$\rightsquigarrow$ $(\mu_t)_{t\geq}$ is a *Wasserstein Gradient Flow* of F_τ.

$\rightsquigarrow$ Let us study the convergence of this dynamics

Log-Sobolev inequality

Relative Entropy: $H(\mu|\nu) := \int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\mu$

Relative Fisher Information: $I(\mu|\nu) := \int \|\nabla \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\|^2 \,\mathrm{d}\mu$

Main assumption: Log-Sobolev Inequality

Assume that there exists $\rho_\tau > 0$ such that $\forall \mu \in \mathcal{P}_2(\mathcal{X})$, the probability measure $\nu \propto e^{-V[\mu]/\tau}$ satisfies LSI(ρ_τ) i.e.

$$H(\tilde{\mu}|\nu) \leq \frac{1}{2\rho_\tau} I(\tilde{\mu}|\nu), \quad \forall \tilde{\mu} \in \mathcal{P}(\mathcal{X}).$$

LSI = Łojasiewicz inequality for $\mu \mapsto H(\mu|\nu)$ in Wasserstein space.

It is satisfied:

- if $\mathcal{X}$ is compact and $\|V[\mu]\|_\infty < \infty$ (uniformly in μ)
- if $\mathcal{X} = \mathbb{R}^d$ and $\|V[\mu] - f\|_\infty < \infty$ for some f strongly convex.

Exponential convergence

Theorem [Nitanda et al. 2022], [Chizat 2022]

Under the previous assumptions the Mean-Field Langevin dynamics is well-posed and converges globally at an exponential rate:

$$F_\tau(\mu_t) - F_\tau(\mu^*_\tau) \leq e^{-2\tau\rho_\tau t}\big(F_\tau(\mu_0) - F_\tau(\mu^*_\tau)\big).$$

The same rate holds for $W_2^2(\mu_t, \mu^*_\tau)$ and $H(\mu_t|\mu^*_\tau)$.

- recovers the rate for Langevin in the linear case

$$G(\mu) = \int V \,\mathrm{d}\mu$$

- also a convergence speed for simulated annealing (see paper)

Nitanda, Wu, Suzuki (2022). *Convex Analysis of the Mean Field Langevin Dynamics.*
Chizat (2022). *Mean-Field Langevin Dynamics: Exponential Convergence and Annealing.*

Proof Idea ($\tau = 1$)

Proof: the standard linear case (Langevin).

Let $V \in \mathcal{C}^2(\mathcal{X})$, $\nu \propto e^{-V}$ and $F(\mu) := \int V \, \mathrm{d}\mu + H(\mu) = H(\mu|\nu)$.
By direct computations and Log-Sobolev inequality:

$$\frac{\mathrm{d}}{\mathrm{d}t}(F(\mu_t) - F^*) = -I(\mu_t|\nu) \leq -2\rho_\tau H(\mu_t|\nu) = -2\rho_\tau (F(\mu_t) - F^*).$$

Proof Idea ($\tau = 1$)

Proof: the standard linear case (Langevin).

Let $V \in \mathcal{C}^2(\mathcal{X})$, $\nu \propto e^{-V}$ and $F(\mu) := \int V \, \mathrm{d}\mu + H(\mu) = H(\mu|\nu)$.
By direct computations and Log-Sobolev inequality:

$$\frac{\mathrm{d}}{\mathrm{d}t}(F(\mu_t) - F^*) = -I(\mu_t|\nu) \leq -2\rho_\tau H(\mu_t|\nu) = -2\rho_\tau (F(\mu_t) - F^*).$$

Proof: the general case

Defining $\nu_t \propto e^{-V[\mu_t]}$, it holds:

Energy Dissipation Ineq.

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\mu_t) = -I(\mu_t|\nu_t)$$

Entropy Sandwich Lemma

$$H(\mu_t|\mu^*) \leq F(\mu_t) - F^* \leq H(\mu_t|\nu_t).$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(F(\mu_t) - F^*) = -I(\mu_t|\nu_t) \leq -2\rho_\tau H(\mu_t|\nu_t) \leq -2\rho_\tau (F(\mu_t) - F^*).$$

Prior works & Applications

Prior works:

- Mei, Montanari, Nguyen (2018). *A Mean Field View of the Landscape of Two-Layers Neural Networks.*
- Hu, Ren, Siska, Szpruch (2019). *Mean-Field Langevin Dynamics and Energy Landscape of Neural Networks.*
- Kazeykina, Ren, Tan, Yang (2020). *Ergodicity of the underdamped Mean-Field Langevin dynamics.*

Applications:

- Noisy Gradient Descent on wide two-layer neural networks
- Free-support debiaised entropic Wasserstein barycenters (Chizat, in prep)
- Min-entropy estimator for trajectory inference (Chizat, Zhang, Heitz, Schiebinger, 2022)

Back to Trajectory Inference

Reminders

Path-space formulation over $\mathcal{P}(\Omega)$:

$$\mathcal{F}(R) := \mathrm{Fit}(R_{t_1}, \dots, R_{t_T}) + \tau H(R|W^\tau)$$

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

$$F(\mu) := \underbrace{\mathrm{Fit}(\mu_1, \dots, \mu_T) + \sum_{i=1}^{T-1} \frac{1}{\Delta t_i} T_{\tau \Delta t_i}(\mu_i, \mu_{i+1})}_{G(\mu)} + \tau \underbrace{\sum_{i=1}^{T} H(\mu_i)}_{H(\mu)}$$

Theorem

There is a computable bijection between minimizers of $\mathcal{F}$ and F.

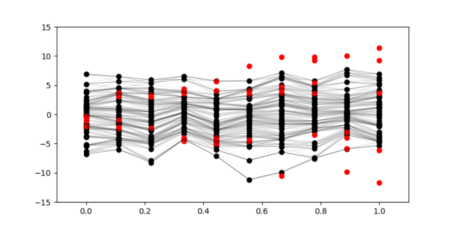
- G is not convex but $G + \tau H$ is
- Apply MFL to $F_\epsilon = G + (\tau + \epsilon) H$ for some $\epsilon > 0$

In (Chizat, Zhang, Heitz, Schiebinger, 2022)
Adapted from (Benamou et al. 2019), (Lavenant et al. 2020)

Theory & Practice: trajectory inference via MFL

Theorem

If $\mathcal{X}$ is compact, the Mean-Field Langevin dynamics $(\mu_s)_{s\geq 0}$ for F_ϵ is well-posed and converges exponentially to minimizers of F_ϵ at a rate $e^{-K/\epsilon}$.
With $\epsilon(s) = C/\log s$ one has $F_0(\mu_s) - \inf F_0 \lesssim 1/\log s$.



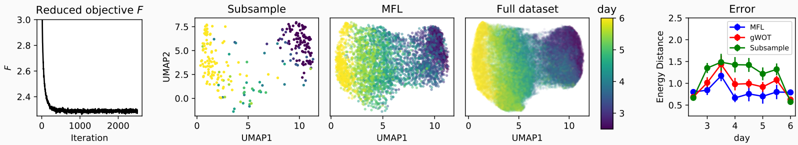
Chizat, Zhang, Heitz, Schiebinger (2022). *Trajectory Inference via Mean-field Langevin in Path Space*.

Theory & Practice: trajectory inference via MFL

Theorem

If $\mathcal{X}$ is compact, the Mean-Field Langevin dynamics $(\boldsymbol{\mu}_s)_{s\geq 0}$ for F_ϵ is well-posed and converges exponentially to minimizers of F_ϵ at a rate $e^{-K/\epsilon}$.

With $\epsilon(s) = C/\log s$ one has $F_0(\boldsymbol{\mu}_s) - \inf F_0 \lesssim 1/\log s$.



In paper: extension to deal with cell birth/death

Chizat, Zhang, Heitz, Schiebinger (2022). *Trajectory Inference via Mean-field Langevin in Path Space*.

Stability of Entropic Optimal Transport

Lipschitz continuity of the Schrödinger map

The first-variation (ϕ, ψ) of T_λ solves the **Schrödinger system**:

$$\begin{cases} \varphi(x) = -\lambda \log \int_{\mathcal{X}} e^{(\psi(y)-c(x,y))/\lambda}\, \mathrm{d}\nu(y) \\ \psi(y) = -\lambda \log \int_{\mathcal{X}} e^{(\phi(x)-c(x,y))/\lambda}\, \mathrm{d}\mu(x) \end{cases}$$

Theorem (Stability of EOT [Carlier, Chizat, Laborde, in prep.])

If $\mathcal{X}$ compact and $c \in \mathcal{C}^k(\mathcal{X} \times \mathcal{X})$ with $k \geq 1$ then $\exists C_k > 0$ s.t.

$$\|(\varphi, \psi) - (\tilde{\varphi}, \tilde{\psi})\|_{\mathcal{C}^k/\sim} \leq C_k\big(W_2(\mu, \tilde{\mu}) + W_2(\nu, \tilde{\nu})\big)$$

- $\mathcal{C}^k/\sim$ is the usual $\mathcal{C}^k$ norm quotiented by the equiv. relation

$$(\varphi, \psi) \sim (\varphi + \kappa, \psi - \kappa), \quad \kappa \in \mathbb{R}$$

- regularizing by $\mu \otimes \nu$ is crucial
- result proved in the multi-marginal case

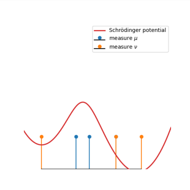
Proof idea: Implicit Function Theorem

Notation: $\boldsymbol{\varphi} \leftarrow (\varphi, \psi), \boldsymbol{\mu} \leftarrow (\mu, \nu)$

- write the Schrödinger system as $F(\boldsymbol{\varphi}, \boldsymbol{\mu}) = 0$
- consider any transport plan $\gamma \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$, the induced interpolation $\boldsymbol{\mu}_t = ((1-t)\pi^1 + t\pi^2)_{\#}\gamma$ and the functional

$$G : \left| \begin{array}{r} (\mathcal{C}^k / \sim) \times [0,1] \to (\mathcal{C}^k / \sim) \\ (\boldsymbol{\varphi}, t) = F(\boldsymbol{\varphi}, \boldsymbol{\mu}_t) \end{array} \right.$$

- apply the Implicit Function Theorem in Banach space



See also: Carlier, Laborde (2020), Nutz and Wiesel (2022)

Some remarks

- **Mean-Field Langevin interacts nicely with Entropic OT**
- **Statistical guarantees for the estimator?**
- **Theory of diffusion in path-space?**

References

- **(Chizat, '22) Mean Field Langevin Dynamics: Exponential convergence and annealing.**
- **(Carlier, Chizat, Laborde, in prep.) Lipschitz continuity of the Schrödinger map in Entropic Optimal Transport.**
- **(Chizat, Zhang, Heitz, Schiebinger, '22) Trajectory Inference via Mean-Field Langevin in Path Space.**