Nonlinear Filtering with Brenier Optimal Transport Maps

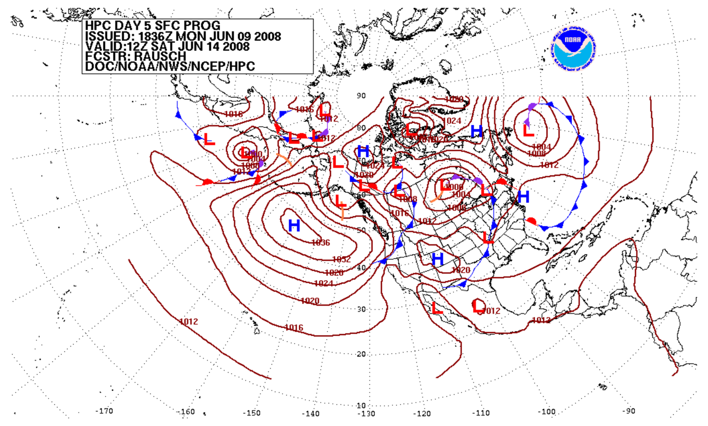
UNIVERSITY of WASHINGTON

Mohammad Al-Jarrah, Niyizhen "Jenny" Jin, Bamdad Hosseini, Amirhossein Taghvaei

The Forty-first International Conference on Machine Learning (ICML), Vienna, Austria, July, 2024

Motivation: Climate forecasting

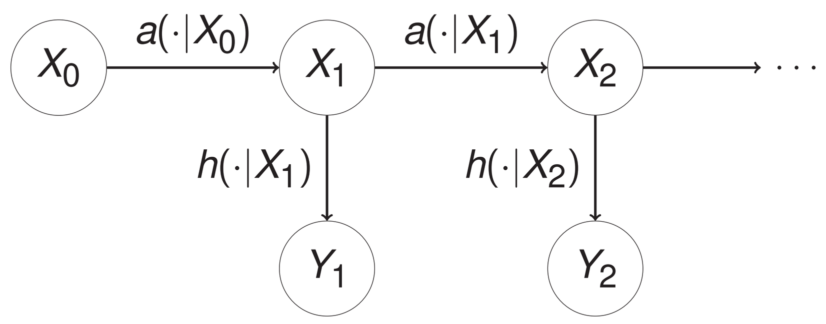
- **Hidden state:** Temperature field of the atmosphere
- **Measurements:** Surface observations from automated weather stations at ground level over land and from weather buoys at sea
- **Problem:** Predict the temperature of the atmosphere for a given location and time.



Mathmatical formulation of the filtering problem

Dynamical system:

- State process: $X_k \in \mathbb{R}^n$
- Observation process: $Y_k \in \mathbb{R}^m$



Problem:

Given: $\{Y_1, \ldots\ldots, Y_k\}$

Find posterior dist.: $\pi_k(\cdot) = P(X_k \in \cdot \mid Y_1, \ldots, Y_k)$

Particle filter and the curse of dimensionality

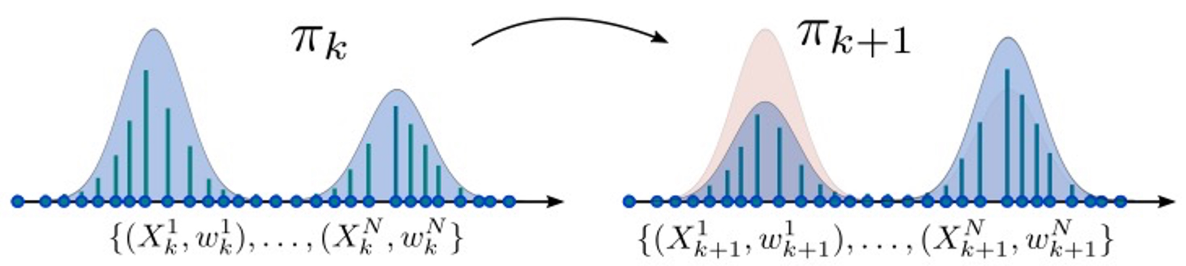
Particle filter methodology

- Approximate π_k with weighted empirical distribution of particles
- Apply the update rule to the particles and weights

$$X^i_{k+1} = a(\cdot|X^i_k), \quad w^i_{k+1} \propto h(Y_{k+1}|X^i_{k+1})$$

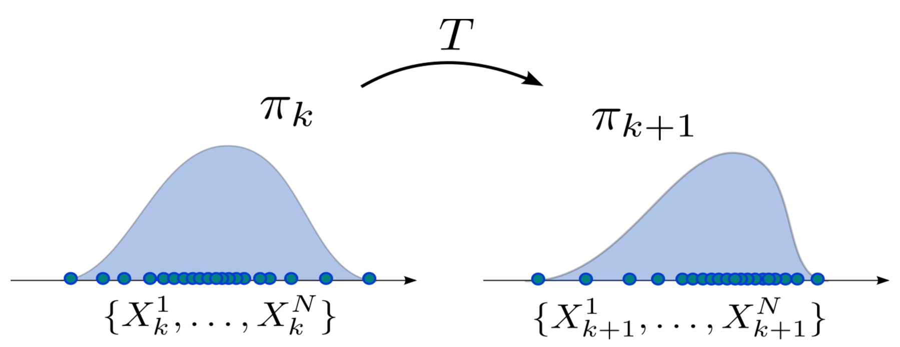
Properties:

- Exact in the limit as number of particles goes to ∞
- Suffer from weight degeneracy as the dimension increases



Transport/coupling view point

Transport approach: update particle with a transport map T from π_k to π_{k+1}



Method: Optimal transport formulation of the Bayes' law

Bayes' Law: $\mathbf{P}_{X|Y} = \frac{\mathbf{P}_X \mathbf{P}_{Y|X}}{\mathbf{P}_Y} = T(\cdot; Y)_{\#}\mathbf{P}_X$

where T is the solution to

$$\max_{f \in c\text{-Concave}_x} \min_{T \in \mathcal{M}(P_X \otimes P_Y)} \mathbb{E}\left[f(X,Y) - f(T(\overline{X},Y),Y) + \frac{1}{2}\|T(\overline{X},Y) - \overline{X}\|^2\right] \quad (*)$$

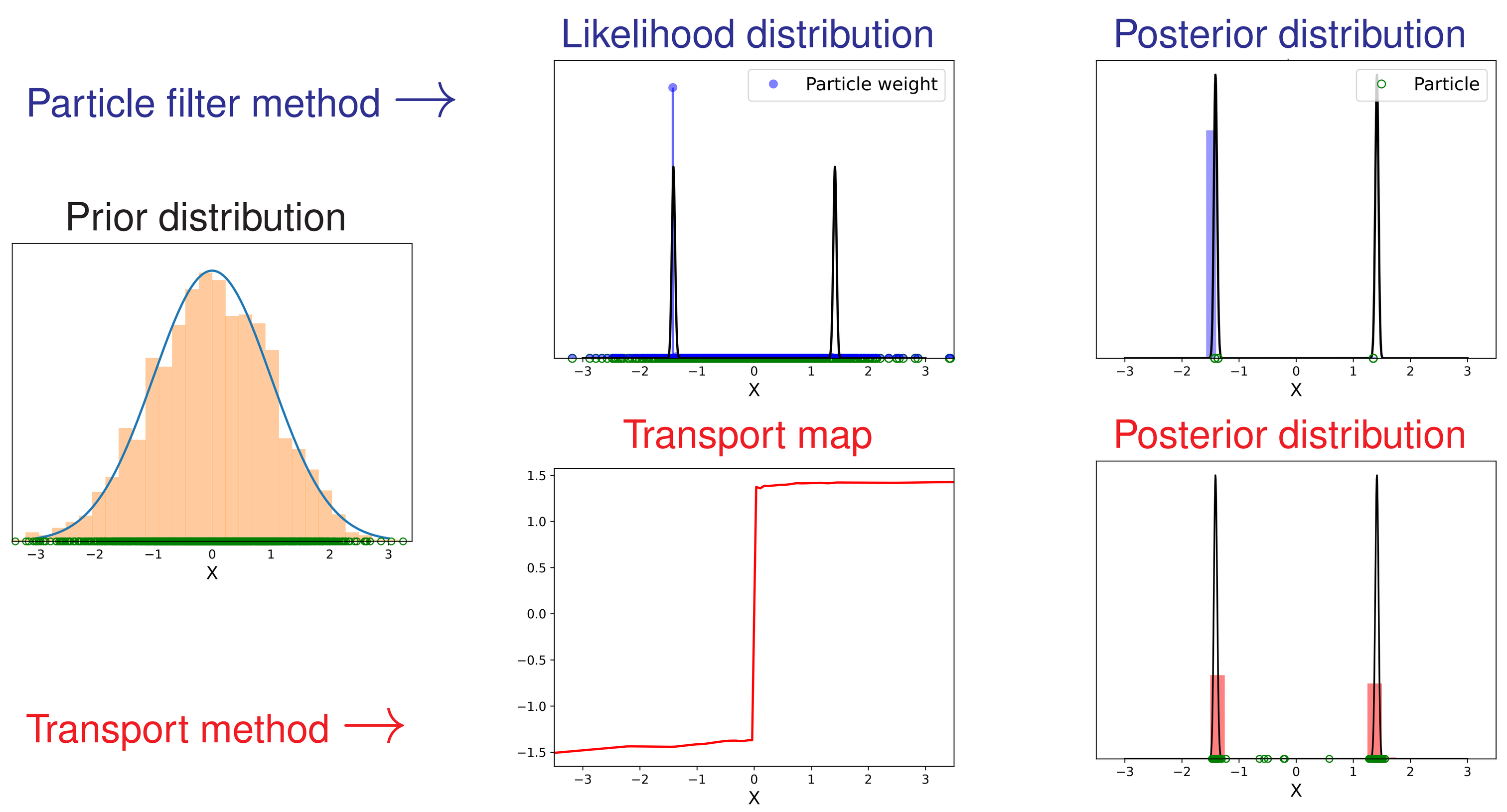
Properties:

- Only requires samples $(X_i, Y_i) \sim \mathbf{P}_{XY}$ (data-driven / simulation-based)
- Enable construction of "approximate" posterior distribution
- Allow application of ML tools (Stochastic optimization and Neural Networks)

Illustrative example with a likelihood degenerate model

$$Y = \frac{1}{2}X^2 + \sqrt{0.04}\cdot W, \quad W \sim \mathcal{N}(0,1)$$

Goal: Compute the conditional distribution of X given Y



Existence and consistency analysis

Assume:

- P_X is absolutely continuous and has a finite second moment
- The posterior $P_{X|Y=y}$ admits a density with respect to the Lebesgue measure $\forall y$

Then:

- There exists a unique pair $(\overline{f}, \overline{T})$ solves $(*)$
- The map $\overline{T}(\cdot, y)$ is the OT map from π to $P_{X|Y=y}$ for a.e. y.

Error analysis

Assume:

- The same assumptions above hold,
- Let (f, T) be a possibly non-optimal pair with an optimality gap $\epsilon(f, T)$.
- Assume $x \mapsto \frac{1}{2}\|x\|^2 - f(x,y)$ is α-strongly convex in x for all y.

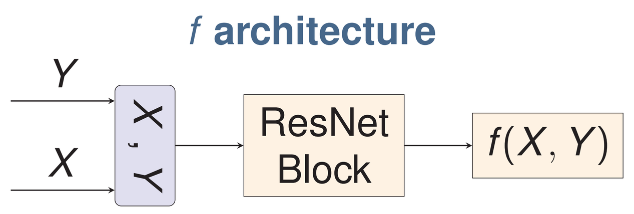
Then:

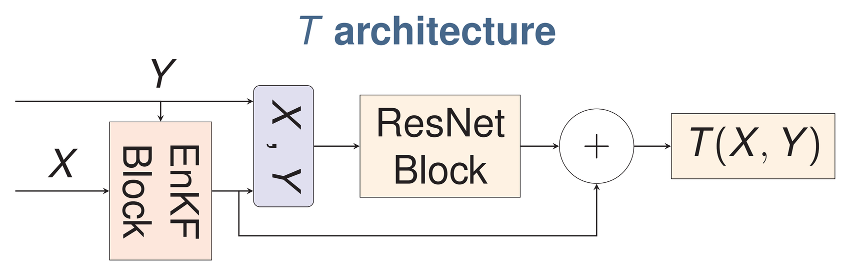
$$\mathbb{E}[\|T(\overline{X},Y) - \overline{T}(\overline{X},Y)\|^2] \leq \frac{4}{\alpha}\epsilon(f,T).$$

Optimal Transport Particle Filter Algorithm

Input: Initial particles $\{X^i_0\}_{i=1}^N$, observation signal $\{Y_t\}_{t=1}^{t_f}$, and $a(x \mid x')$, $h(y \mid x)$ kernels
Initialize: initialize neural net f, T according to block architecture
for $t = 1$ to t_f **do**
  Propagation: $X^i_{t|t-1} \sim a(. \mid X^i_{t-1})$ and $Y^i_t \sim h(. \mid X^i_{t|t-1})$
  Optimization: Update the weight parameters of f, T throughout a gradient ascent-descent procedure for $(*)$
  Conditioning: Update particles $X^i_t = T(X^i_{t|t-1}, Y_t)$, $\quad \forall i = 1, \ldots N$.
end for
Output: Particles $\{X^i_t\}_{i=1}^N$ for $t = 0, \ldots, t_f$.

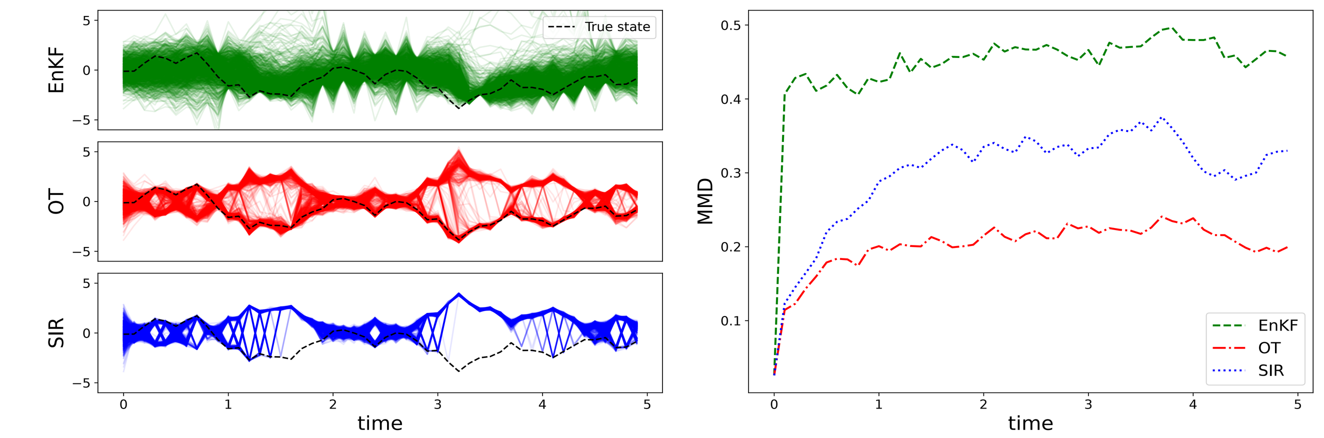
Neural net f, T architectures





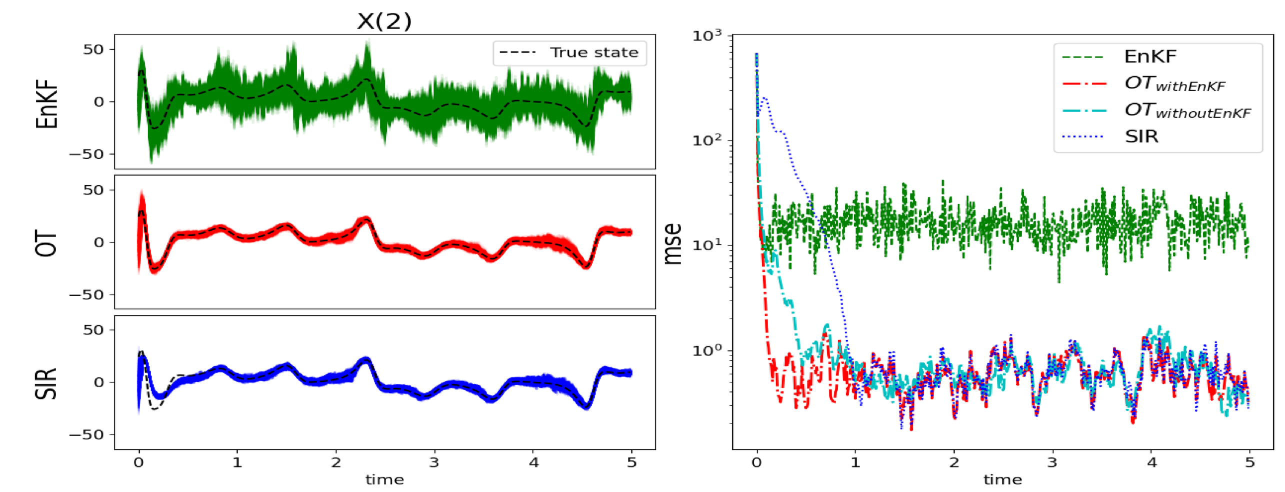
Numerical result: Bimodal dynamic example

$$X_t = 0.9\cdot X_{t-1} + 2\sqrt{0.1}\cdot V_t, \quad X_0 \sim \mathcal{N}(0, I_n)$$
$$Y_t = X_t \odot X_t + \sqrt{0.1}\cdot W_t,$$



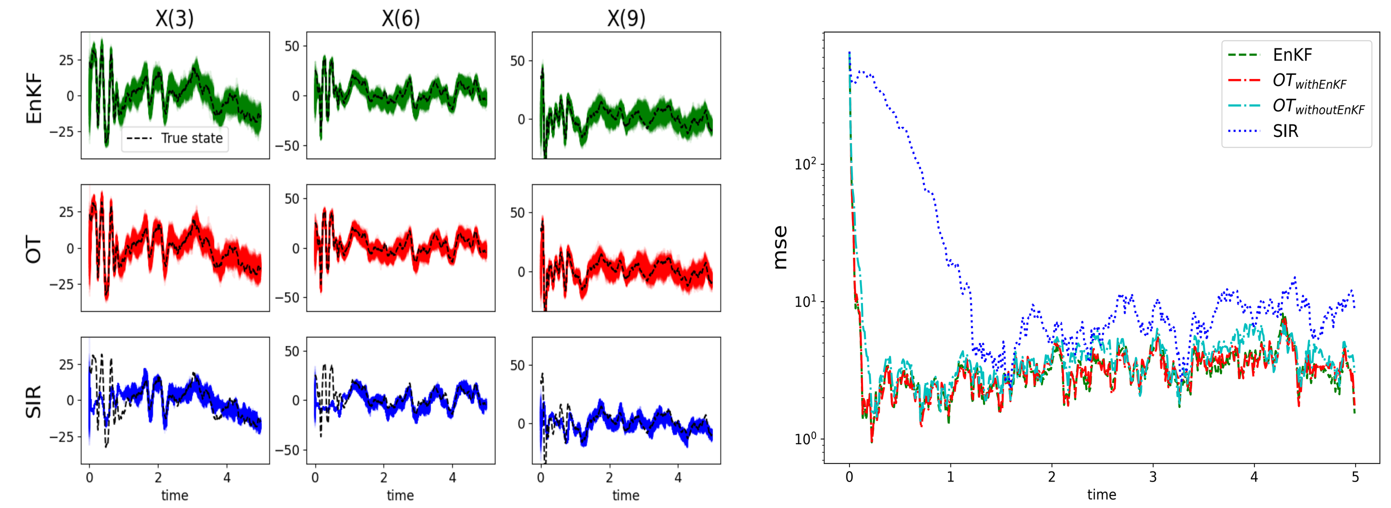
Numerical result: Lorenz 63 model

Dynamical system Lorenz 63 model with observing the first and third states.

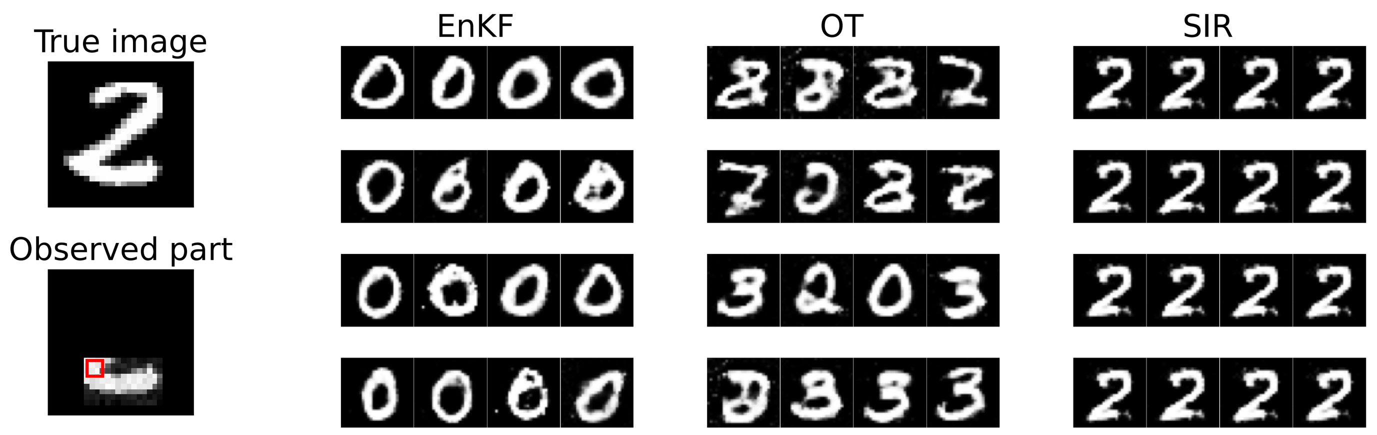


Numerical result: Lorenz 96 model

Dynamical system Lorenz 96 model with observing every other two states.



Numerical result: Image in-painting on MNIST



Future directions of research

- Explore alternative architectures to increase efficiency
- Verification of the algorithm on real-world applications
- Application for decision-making under uncertainty

Acknowledgement and GitHub page

Financial support from the National Science Foundation (NSF) EPCN-2318977, and DMS-2208535 are gratefully acknowledged.