Variational Optimal Transport Methods for Nonlinear Filtering

Presented at the
2024 SIAM Conference on Uncertainty Quantification

Amirhossein Taghvaei
Joint work with Mohammad Al-Jarrah, Jenny Jin, and Bamdad Hosseini

Department of Aeronautics & Astronautics
University of Washington, Seattle

Feb 27, 2024

Outline

- **Part I:** Bayes' law $\rightarrow$ optimal transport maps
- **Part II:** Application to nonlinear filtering

Outline

- **Part I:** Bayes' law $\rightarrow$ optimal transport maps
- **Part II:** Application to nonlinear filtering

Bayes' law

Problem:

- Hidden random variable X
- Observed random variable Y
- What is the conditional probability distribution of X given Y? (posterior)

Bayes' law: $P_{X|Y} = \dfrac{P_X P_{Y|X}}{P_Y}$

Simple to express, but difficult to implement numerically

Bayes' law

Problem:

- Hidden random variable X
- Observed random variable Y
- What is the conditional probability distribution of X given Y? (posterior)

Bayes' law: $$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y}$$

Simple to express, but difficult to implement numerically

Bayes' law

Problem:

- Hidden random variable X
- Observed random variable Y
- What is the conditional probability distribution of X given Y? (posterior)

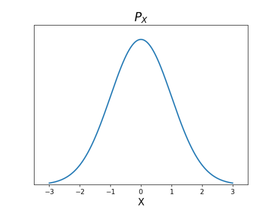
Bayes' law: $$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y}$$

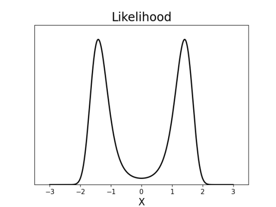
Simple to express, but difficult to implement numerically

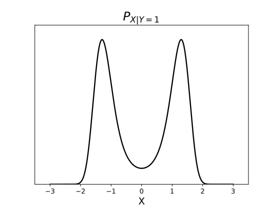
Example:

- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} =?$

Importance sampling (IS):







small noise regime: $\epsilon \to 0$

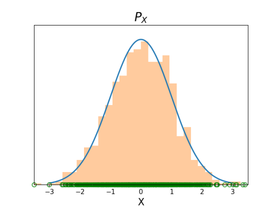
This is the main reason for the curse of dimensionality of IS-based particle filters

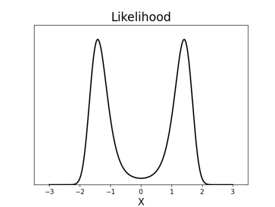
Example:

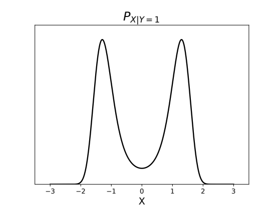
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

- $X^i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$
- $w^i \propto P(Y=1|X^i)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$







small noise regime: $\epsilon \to 0$

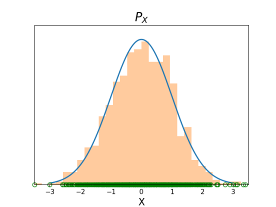
This is the main reason for the curse of dimensionality of IS-based particle filters

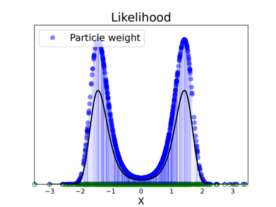
Example:

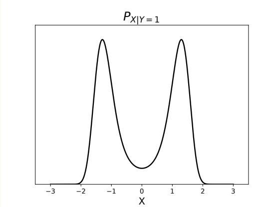
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

- $X^i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$
- $w^i \propto P(Y=1|X^i)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$







small noise regime: $\epsilon \to 0$

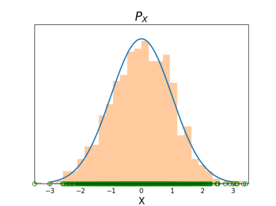
This is the main reason for the curse of dimensionality of IS-based particle filters

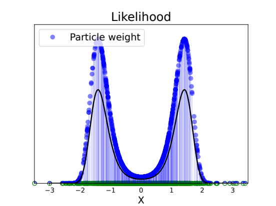
Example:

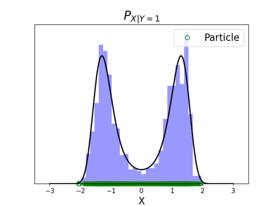
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

- $X^i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$
- $w^i \propto P(Y=1|X^i)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$







small noise regime: $\epsilon \to 0$

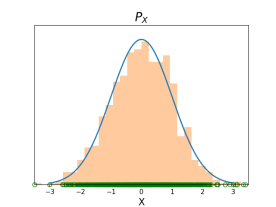
This is the main reason for the curse of dimensionality of IS-based particle filters

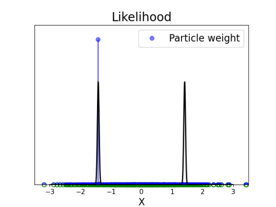
Example:

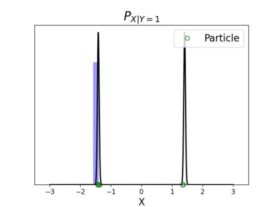
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

- $X^i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$
- $w^i \propto P(Y=1|X^i)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$







small noise regime: $\epsilon \to 0$

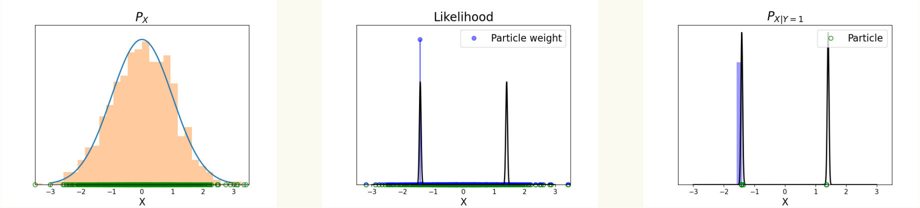
This is the main reason for the curse of dimensionality of IS-based particle filters

Example:

- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

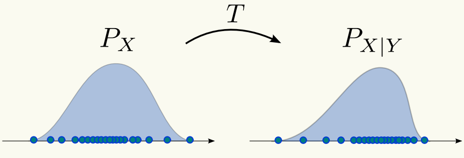
- $X^i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$
- $w^i \propto P(Y=1|X^i)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$



small noise regime: $\epsilon \to 0$

This is the main reason for the curse of dimensionality of IS-based particle filters

Conditioning with transport maps

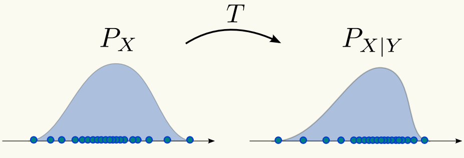


$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

Example:

Questions:

Conditioning with transport maps

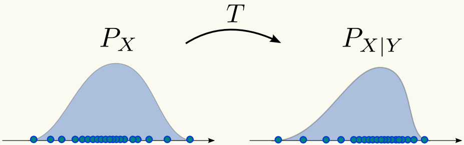


$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

Example:

- Consider $Y = X$. Then, $P_{X|Y=y} = \delta_y$ is represented by the map $T(x,y) = y$
- Consider jointly Gaussian (X, Y). Then $P_{X|Y=y}$ is represented by the (stochastic) map $X \mapsto X + K(y - Y)$

Questions:



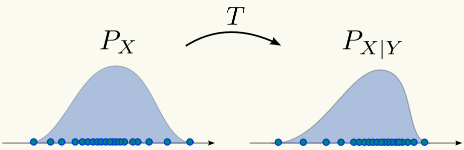
$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

Example:

- Consider $Y = X$. Then, $P_{X|Y=y} = \delta_y$ is represented by the map $T(x, y) = y$
- Consider jointly Gaussian (X, Y). Then $P_{X|Y=y}$ is represented by the (stochastic) map $X \mapsto X + K(y - Y)$

Questions:

Conditioning with transport maps



$$X^i \sim P_X \ \longrightarrow \ T(X^i, y) \sim P_{X|Y=y}$$

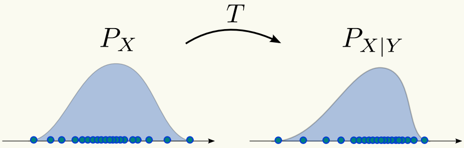
Example:

- Consider $Y = X$. Then, $P_{X|Y=y} = \delta_y$ is represented by the map $T(x,y) = y$
- Consider jointly Gaussian (X,Y). Then $P_{X|Y=y}$ is represented by the (stochastic) map $X \mapsto X + K(y - Y)$

Questions:

- Does the map exists?
- How to numerically find it?

Conditioning with transport maps



$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

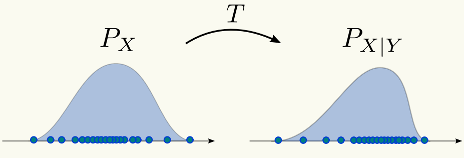
Example:

- Consider $Y = X$. Then, $P_{X|Y=y} = \delta_y$ is represented by the map $T(x, y) = y$
- Consider jointly Gaussian (X, Y). Then $P_{X|Y=y}$ is represented by the (stochastic) map $X \mapsto X + K(y - Y)$

Questions:

- Does the map exists?
- How to numerically find it?

Conditioning with transport maps



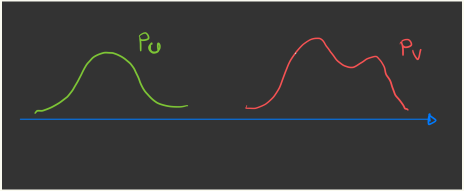
$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

Example:

- Consider $Y = X$. Then, $P_{X|Y=y} = \delta_y$ is represented by the map $T(x, y) = y$
- Consider jointly Gaussian (X, Y). Then $P_{X|Y=y}$ is represented by the (stochastic) map $X \mapsto X + K(y - Y)$

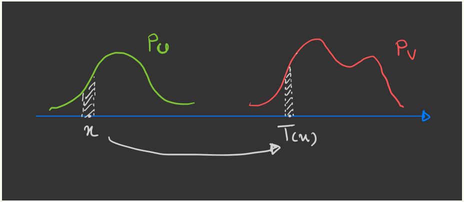
Questions:

- Does the map exists?
- How to numerically find it?



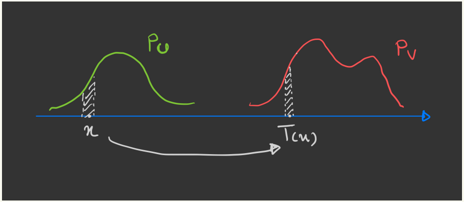
- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_{\#}P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:



- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_\# P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

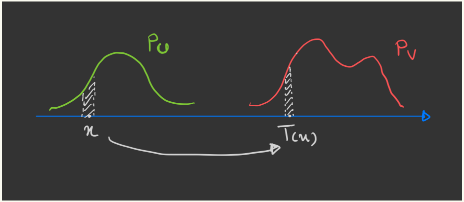
Questions:



- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_{\#}P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:

- Yes, as long as P_U admits Lebesgue density
- max-min stochastic optimization

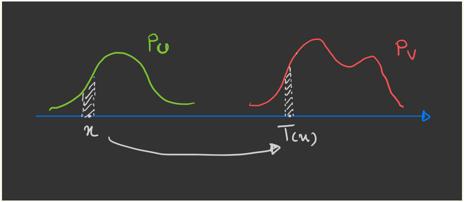


- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_{\#}P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:

- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? max-min stochastic optimization

$$(\overline{f}, \overline{T}) \leftarrow \max_{f \in c\text{-concave}} \min_{T} \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 + f(V) - f(T(U))\right]$$

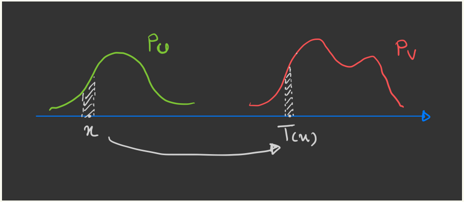


- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_\# P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:

- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? max-min stochastic optimization

$$(\overline{f}, \overline{T}) \leftarrow \max_{f \in c\text{-concave}} \min_{T} \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 + f(V) - f(T(U))\right]$$



- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_\# P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

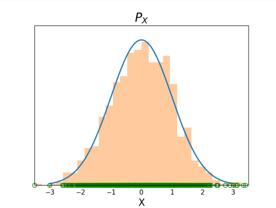
Questions:

- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? max-min stochastic optimization

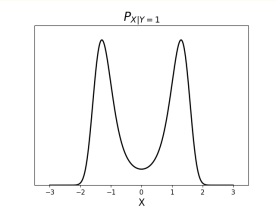
$$(\overline{f}, \overline{T}) \leftarrow \max_{f \in c\text{-concave}} \min_{T} \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 + f(V) - f(T(U))\right]$$

Conditioning with optimal transport map
Illustrative example

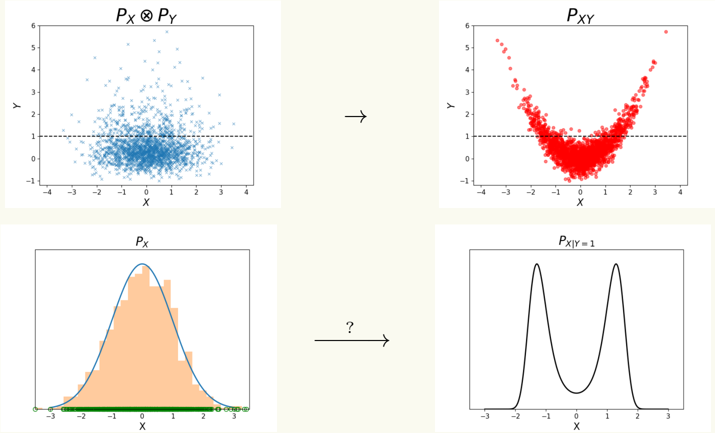
$\rightarrow$



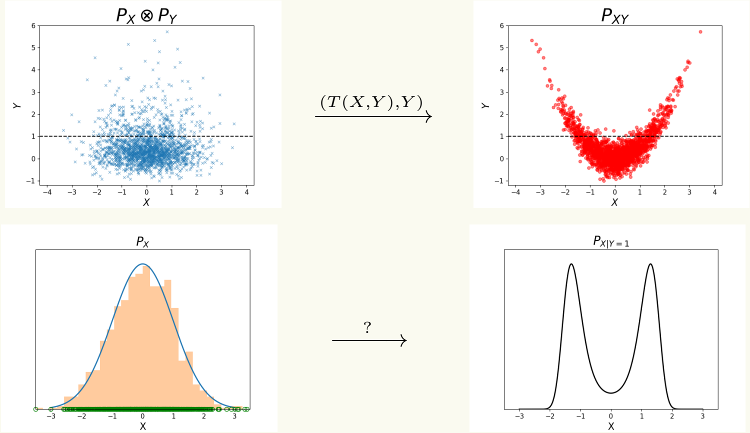
$\xrightarrow{?}$



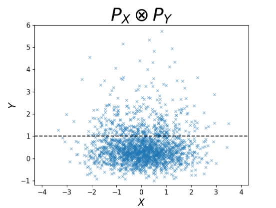
Conditioning with optimal transport map
Illustrative example



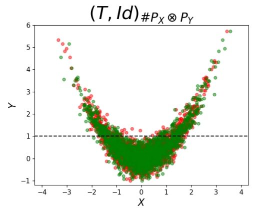
Conditioning with optimal transport map
Illustrative example

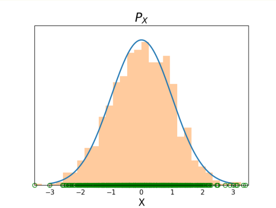


Conditioning with optimal transport map
Illustrative example

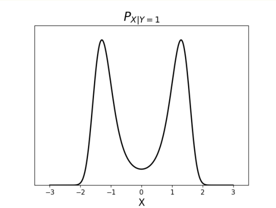


$\xrightarrow{(T(X,Y),Y)}$

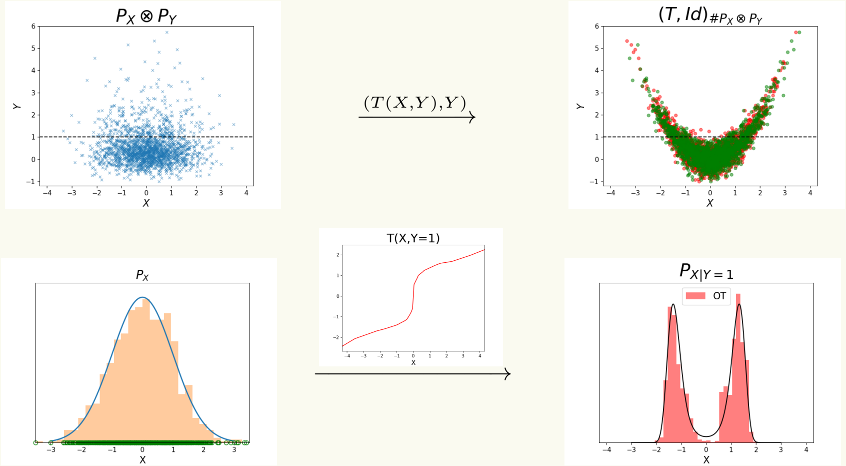




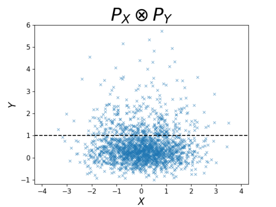
$\xrightarrow{?}$



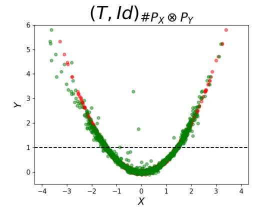
Conditioning with optimal transport map
Illustrative example

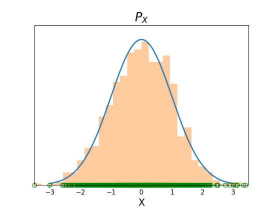


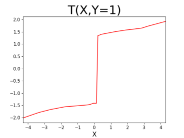
Conditioning with optimal transport map
Illustrative example

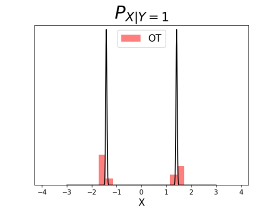


$\xrightarrow{(T(X,Y),Y)}$









small noise limit

Bayes law: $$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y}$$

$$= T(\cdot; Y)\#P_X$$

Conditional max-min formulation:

$$\max_{f \in c\text{-concave}_x} \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X}, Y) - \bar{X}\|^2 + f(X; Y) - f(T(\bar{X}, Y), Y)\right]$$

Computational properties:

- Only requires samples $(X_i, Y_i) \sim P_{XY}$ (data-driven/simulation based)
- Enables construction of "approximate" posterior distributions
- Allows application of ML tools (stochastic optimization and neural nets)

Conditioning with optimal transport map
Variational formulation of the Bayes' law

Bayes law:

$$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y} = \color{red}{T}(\cdot;Y)\#P_X$$

Conditional max-min formulation:

$$\max_{f \in c\text{-concave}_x} \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X},Y) - \bar{X}\|^2 + f(X;Y) - f(T(\bar{X},Y),Y)\right]$$

Computational properties:

- Only requires samples $(X_i, Y_i) \sim P_{XY}$ (data-driven/simulation based)
- Enables construction of "approximate" posterior distributions
- Allows application of ML tools (stochastic optimization and neural nets)

Conditioning with optimal transport map

Variational formulation of the Bayes' law

Bayes law: $$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y}$$
$$= T(\cdot; Y) \# P_X$$

Conditional max-min formulation:

$$\max_{f \in c\text{-concave}_x} \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X}, Y) - \bar{X}\|^2 + f(X; Y) - f(T(\bar{X}, Y), Y)\right]$$

Computational properties:

- Only requires samples $(X_i, Y_i) \sim P_{XY}$ (data-driven/simulation based)
- Enables construction of "approximate" posterior distributions
- Allows application of ML tools (stochastic optimization and neural nets)

Bayes law: $$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y} = T(\cdot;Y)\#P_X$$

Conditional max-min formulation:

$$\max_{f \in c\text{-concave}_x} \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X},Y)-\bar{X}\|^2 + f(X;Y) - f(T(\bar{X},Y),Y)\right]$$

Computational properties:

- Only requires samples $(X_i, Y_i) \sim P_{XY}$ (data-driven/simulation based)
- Enables construction of "approximate" posterior distributions
- Allows application of ML tools (stochastic optimization and neural nets)

- Variational problem: $\max_{f} \min_{T} J(f, T; P_{X,Y})$
- max-min optimality gap: $\epsilon(f, T)$

(Conditional) Brenier's theorem

- (Well-posedness) If P_X admits (Lebesgue) density, then, there exists a unique pair $(\overline{f}, \overline{T})$ that solves the variational problem and

$$\overline{T}(\cdot, y)\#P_X = P_{X|Y=y}, \quad \text{a.e} \quad y$$

- (Sensitivity) Let (f, T) be a possibly non-optimal pair. Assume $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y)$ is α-strongly convex for all y. Then,

$$d(T(\cdot, Y)\#P_X, P_{X|Y}) \leq \sqrt{\frac{4}{\alpha}\epsilon(f, T)}.$$

Carlier, G., Chernozhukov, V., and Galichon, A. (2016). Vector quantile regression: an optimal trans- port approach. The Annals of Statistics,

- Variational problem: $\max_f \min_T J(f,T;P_{X,Y})$
- max-min optimality gap: $\epsilon(f,T)$

(Conditional) Brenier's theorem

- (Well-posedness) If P_X admits (Lebesgue) density, then, there exists a unique pair $(\overline{f},\overline{T})$ that solves the variational problem and

$$\overline{T}(\cdot,y)\#P_X = P_{X|Y=y}, \quad \text{a.e} \quad y$$

- (Sensitivity) Let (f,T) be a possibly non-optimal pair. Assume $x \mapsto \frac{1}{2}\|x\|^2 - f(x,y)$ is α-strongly convex for all y. Then,

$$d(T(\cdot,Y)\#P_X, P_{X|Y}) \leq \sqrt{\frac{4}{\alpha}\epsilon(f,T)}.$$

Carlier, G., Chernozhukov, V., and Galichon, A. (2016). Vector quantile regression: an optimal trans- port approach. The Annals of Statistics,

- Variational problem: $\max_f \min_T J(f, T; P_{X,Y})$
- max-min optimality gap: $\epsilon(f, T)$

(Conditional) Brenier's theorem

- (Well-posedness) If P_X admits (Lebesgue) density, then, there exists a unique pair $(\overline{f}, \overline{T})$ that solves the variational problem and

$$\overline{T}(\cdot, y)\#P_X = P_{X|Y=y}, \quad \text{a.e} \quad y$$

- (Sensitivity) Let (f, T) be a possibly non-optimal pair. Assume $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y)$ is α-strongly convex for all y. Then,

$$d(T(\cdot, Y)\#P_X, P_{X|Y}) \leq \sqrt{\frac{4}{\alpha}\epsilon(f, T)}.$$

Carlier, G., Chernozhukov, V., and Galichon, A. (2016). Vector quantile regression: an optimal trans- port approach. The Annals of Statistics,

Outline

- **Part I:** Bayes' law $\rightarrow$ optimal transport maps
- **Part II:** Application to nonlinear filtering

Outline

- **Part I:** Bayes' law $\rightarrow$ optimal transport maps
- **Part II:** Application to nonlinear filtering

Nonlinear filtering problem

$$\begin{array}{ccccccc} \longrightarrow & X_{k-1} & \longrightarrow & X_k & \longrightarrow & X_{k+1} & \longrightarrow \\ & \downarrow & & \downarrow & & \downarrow & \\ & Y_{k-1} & & Y_k & & Y_{k+1} & \end{array}$$

- X_t is the state (unknown)
- Y_t is the observation

Questions: Given history of observation $Y_{1:t} := \{Y_1, \ldots, Y_t\}$,

- What is the most likely value of X_t?
- What is the probability of $X_t \in A$?
- What is the best m.s.e estimate for X_t?
- ...

Answer: given by the conditional distribution $\pi_t = P_{X_t|Y_{1:t}}$ (posterior, belief)

Nonlinear filtering: numerical approximation of the posterior π_t for all t.

Nonlinear filtering problem

$$
\begin{tikzcd}
{} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\
 & Y_{k-1} & Y_k & Y_{k+1} &
\end{tikzcd}
$$

- X_t is the state (unknown)
- Y_t is the observation

Questions: Given history of observation $Y_{1:t} := \{Y_1, \ldots, Y_t\}$,

- What is the most likely value of X_t?
- What is the probability of $X_t \in A$?
- What is the best m.s.e estimate for X_t?
- ...

Answer: given by the conditional distribution $\pi_t = P_{X_t|Y_{1:t}}$ (posterior, belief)

Nonlinear filtering: numerical approximation of the posterior π_t for all t.

Nonlinear filtering problem

$$\begin{tikzcd} {} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\ & Y_{k-1} & Y_k & Y_{k+1} & \end{tikzcd}$$

- X_t is the state (unknown)
- Y_t is the observation

Questions: Given history of observation $Y_{1:t} := \{Y_1, \ldots, Y_t\}$,

- What is the most likely value of X_t?
- What is the probability of $X_t \in A$?
- What is the best m.s.e estimate for X_t?
- ...

Answer: given by the conditional distribution $\pi_t = P_{X_t|Y_{1:t}}$ (posterior, belief)

Nonlinear filtering: numerical approximation of the posterior π_t for all t.

Nonlinear filtering problem

$$\begin{tikzcd} {} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\ & Y_{k-1} & Y_k & Y_{k+1} & \end{tikzcd}$$

- X_t is the state (unknown)
- Y_t is the observation

Questions: Given history of observation $Y_{1:t} := \{Y_1, \ldots, Y_t\}$,

- What is the most likely value of X_t?
- What is the probability of $X_t \in A$?
- What is the best m.s.e estimate for X_t?
- ...

Answer: given by the conditional distribution $\pi_t = P_{X_t|Y_{1:t}}$ (posterior, belief)

Nonlinear filtering: numerical approximation of the posterior π_t for all t.

Filtering equations

- $\pi_t := \mathsf{P}(X_t|Y_{1:t})$
- Two important operations:

$$\text{Propagation:} \quad \pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi$$

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} B_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} B_{Y_t}(\mathcal{A}\pi_{t-1}) =: \mathcal{T}_{t,t-1}(\pi_{t-1})$$

- (Exponential) filter stability : $\exists\lambda \in (0,1)$ s.t.

$$d(\mathcal{T}_{t,0}(\pi_0), \mathcal{T}_{t,0}(\tilde{\pi}_0)) \leq C\lambda^k d(\pi_0, \tilde{\pi}_0), \quad \forall \pi_0, \tilde{\pi}_0.$$

Filtering equations

- $\pi_t := \mathsf{P}(X_t | Y_{1:t})$
- Two important operations:

$$\text{Propagation:} \quad \pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi$$

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} B_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} B_{Y_t}(\mathcal{A}\pi_{t-1}) =: \mathcal{T}_{t,t-1}(\pi_{t-1})$$

- (Exponential) filter stability : $\exists \lambda \in (0,1)$ s.t.

$$d(\mathcal{T}_{t,0}(\pi_0), \mathcal{T}_{t,0}(\tilde{\pi}_0)) \leq C\lambda^k d(\pi_0, \tilde{\pi}_0), \quad \forall \pi_0, \tilde{\pi}_0.$$

Filtering equations

- $\pi_t := \mathsf{P}(X_t|Y_{1:t})$
- Two important operations:

$$\text{Propagation:} \quad \pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi$$

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} B_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} B_{Y_t}(\mathcal{A}\pi_{t-1}) =: \mathcal{T}_{t,t-1}(\pi_{t-1})$$

- (Exponential) filter stability : $\exists \lambda \in (0,1)$ s.t.

$$d(\mathcal{T}_{t,0}(\pi_0), \mathcal{T}_{t,0}(\tilde{\pi}_0)) \le C\lambda^k d(\pi_0, \tilde{\pi}_0), \quad \forall \pi_0, \tilde{\pi}_0.$$

Filtering equations

- $\pi_t := \mathsf{P}(X_t|Y_{1:t})$
- Two important operations:

$$\text{Propagation:} \quad \pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi$$

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} B_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} B_{Y_t}(\mathcal{A}\pi_{t-1}) =: \mathcal{T}_{t,t-1}(\pi_{t-1})$$

- (Exponential) filter stability : $\exists \lambda \in (0,1)$ s.t.

$$d(\mathcal{T}_{t,0}(\pi_0), \mathcal{T}_{t,0}(\tilde{\pi}_0)) \leq C\lambda^k d(\pi_0, \tilde{\pi}_0), \quad \forall \pi_0, \tilde{\pi}_0.$$

Optimal Transport Filter
No dynamics setting (for simplicity)

Filter design steps:

$$\text{exact posterior:} \quad \pi_t = \mathcal{B}_{Y_t}(\pi_{t-1})$$

$$\text{mean-field process:} \quad \bar{X}_t = \overline{T}_t(\bar{X}_{t-1}, Y_t)$$

$$\text{particle system:} \quad X_t^i = \hat{T}_t(X_{t-1}^i, Y_t)$$

Variational problem:

$$\leftarrow \max_{f} \min_{T} J(f, T; P)$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J\Big(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_t^i, Y_t^i)}\Big)$$

Posterior approximation:

$$\pi_t \approx \hat{\pi}_t^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

Optimal Transport Filter
No dynamics setting (for simplicity)

Filter design steps:

exact posterior: $\pi_t = \overline{T}_t(\cdot, Y_t)_{\#}\pi_{t-1}$

mean-field process: $\bar{X}_t = \overline{T}_t(\bar{X}_{t-1}, Y_t)$

particle system: $X_t^i = \hat{T}_t(X_{t-1}^i, Y_t)$

Variational problem:

$$\overline{T}_t \leftarrow \max_f \min_T J(f, T; P_{X_t, Y_t})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_t^i, Y_t^i)})$$

Posterior approximation:

$$\pi_t \approx \hat{\pi}_t^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

Filter design steps:

exact posterior: $\pi_t = \overline{T}_t(\cdot, Y_t)_{\#}\pi_{t-1}$

mean-field process: $\bar{X}_t = \overline{T}_t(\bar{X}_{t-1}, Y_t)$

particle system: $X_t^i = \hat{T}_t(X_{t-1}^i, Y_t)$

Variational problem:

$$\overline{T}_t \leftarrow \max_f \min_T J(f, T; P_{X_t, Y_t})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_t^i, Y_t^i)})$$

Posterior approximation:

$$\pi_t \approx \hat{\pi}_t^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

Filter design steps:

$$\text{exact posterior:} \quad \pi_t = \overline{T}_t(\cdot, Y_t)_{\#}\pi_{t-1}$$

$$\text{mean-field process:} \quad \bar{X}_t = \overline{T}_t(\bar{X}_{t-1}, Y_t)$$

$$\text{particle system:} \quad X_t^i = \hat{T}_t(X_{t-1}^i, Y_t)$$

Variational problem:

$$\overline{T}_t \leftarrow \max_f \min_T \; J(f, T; P_{X_t, Y_t})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} \; J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_t^i, Y_t^i)})$$

Posterior approximation:

$$\pi_t \approx \hat{\pi}_t^{(N)} = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

Theorem

Assume

1. The exact filter is exponentially stable
2. Uniform bound $\epsilon_{\mathcal{F},\mathcal{T},N}$ on the max-min optimality gap
3. The function $x \mapsto \frac{1}{2}\|x\|^2 - \hat{f}_t(x,y)$ are α-strongly convex for all t and y.
4. Particles are resampled at each step

Then,

$$d(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t) \leq C\left(\sqrt{\frac{4}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap has bias-variance decomposition

$$\epsilon_{\mathcal{F},\mathcal{T},N} \leq \underbrace{\epsilon_{\mathcal{F},\mathcal{T}}}_{\text{approx. theory}} + \underbrace{\frac{C_{\mathcal{F}}}{\sqrt{N}}}_{\text{statistical generalization}}$$

Theorem

Assume

1. The exact filter is exponentially stable
2. Uniform bound $\epsilon_{\mathcal{F},\mathcal{T},N}$ on the max-min optimality gap
3. The function $x \mapsto \frac{1}{2}\|x\|^2 - \hat{f}_t(x,y)$ are α-strongly convex for all t and y.
4. Particles are resampled at each step

Then,

$$d\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t\Big) \leq C\left(\sqrt{\frac{4}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap has bias-variance decomposition

$$\epsilon_{\mathcal{F},\mathcal{T},N} \leq \underbrace{\epsilon_{\mathcal{F},\mathcal{T}}}_{\text{approx. theory}} + \underbrace{\frac{C_{\mathcal{F}}}{\sqrt{N}}}_{\text{statistical generalization}}$$

Theorem

Assume

1. The exact filter is exponentially stable
2. Uniform bound $\epsilon_{\mathcal{F},\mathcal{T},N}$ on the max-min optimality gap
3. The function $x \mapsto \frac{1}{2}\|x\|^2 - \hat{f}_t(x,y)$ are α-strongly convex for all t and y.
4. Particles are resampled at each step

Then,

$$d(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t) \leq C\left(\sqrt{\frac{4}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap has bias-variance decomposition

$$\epsilon_{\mathcal{F},\mathcal{T},N} \leq \underbrace{\epsilon_{\mathcal{F},\mathcal{T}}}_{\text{approx. theory}} + \underbrace{\frac{C_{\mathcal{F}}}{\sqrt{N}}}_{\text{statistical generalization}}$$

Theorem

Assume

1. The exact filter is exponentially stable
2. Uniform bound $\epsilon_{\mathcal{F},\mathcal{T},N}$ on the max-min optimality gap
3. The function $x \mapsto \frac{1}{2}\|x\|^2 - \hat{f}_t(x,y)$ are α-strongly convex for all t and y.
4. Particles are resampled at each step

Then,

$$d\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t\right) \leq C\left(\sqrt{\frac{4}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap has bias-variance decomposition

$$\epsilon_{\mathcal{F},\mathcal{T},N} \leq \underbrace{\epsilon_{\mathcal{F},\mathcal{T}}}_{\text{approx. theory}} + \underbrace{\frac{C_{\mathcal{F}}}{\sqrt{N}}}_{\text{statistical generalization}}$$

Theorem

Assume

1. The exact filter is exponentially stable
2. Uniform bound $\epsilon_{\mathcal{F},\mathcal{T},N}$ on the max-min optimality gap
3. The function $x \mapsto \frac{1}{2}\|x\|^2 - \hat{f}_t(x,y)$ are α-strongly convex for all t and y.
4. Particles are resampled at each step

Then,

$$d\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}, \pi_t\Big) \leq C\left(\sqrt{\frac{4}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap has bias-variance decomposition

$$\epsilon_{\mathcal{F},\mathcal{T},N} \leq \underbrace{\epsilon_{\mathcal{F},\mathcal{T}}}_{\text{approx. theory}} + \underbrace{\frac{C_{\mathcal{F}}}{\sqrt{N}}}_{\text{statistical generalization}}$$

Theorem

Assume

1. The exact filter is exponentially stable
2. Uniform bound $\epsilon_{\mathcal{F},\mathcal{T},N}$ on the max-min optimality gap
3. The function $x \mapsto \frac{1}{2}\|x\|^2 - \hat{f}_t(x,y)$ are α-strongly convex for all t and y.
4. Particles are resampled at each step

Then,

$$d\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t\right) \leq C\left(\sqrt{\frac{4}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap has bias-variance decomposition

$$\epsilon_{\mathcal{F},\mathcal{T},N} \leq \underbrace{\epsilon_{\mathcal{F},\mathcal{T}}}_{\text{approx. theory}} + \underbrace{\frac{C_{\mathcal{F}}}{\sqrt{N}}}_{\text{statistical generalization}}$$

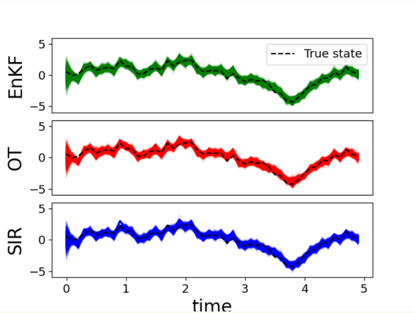
Optimal Transport Filter
Numerical example

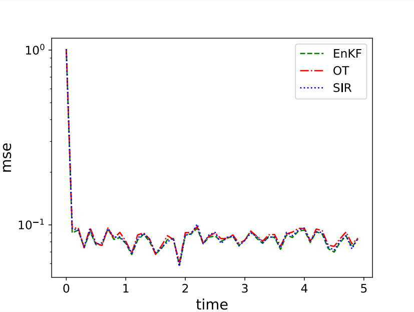
$$
\begin{aligned}
X_t &= (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),\\
Y_t &= X_t + \sigma_W W_t,
\end{aligned}
$$

- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)

Optimal Transport Filter
Numerical example

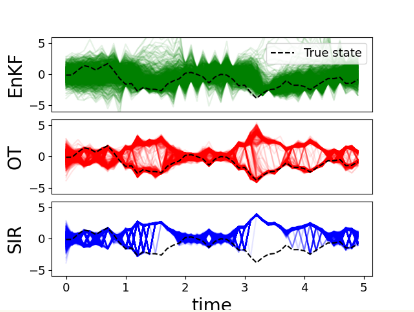
$$X_t = (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),$$
$$Y_t = X_t + \sigma_W W_t,$$

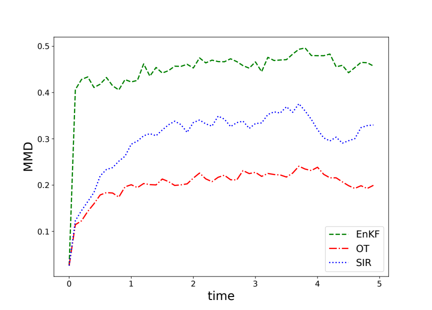




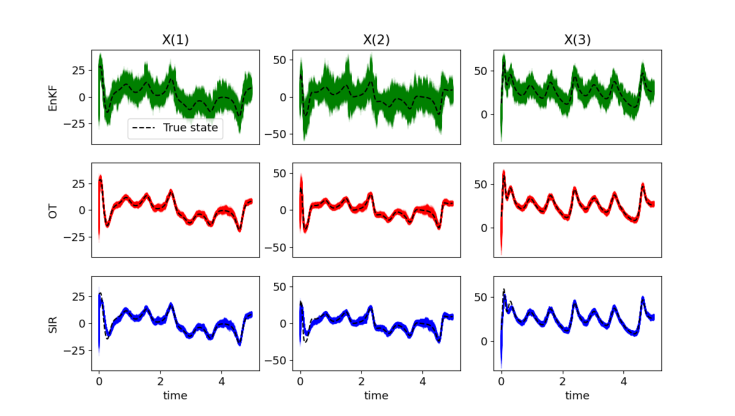
- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)

$$X_t = (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),$$
$$Y_t = X_t^2 + \sigma_W W_t,$$





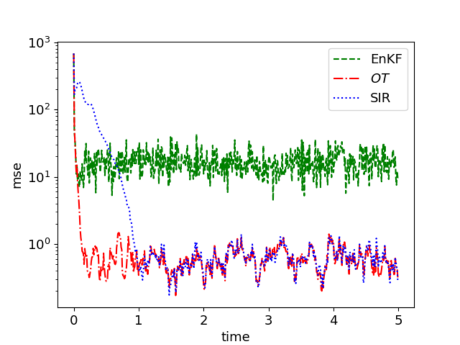
- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)



- Trajectory of the particles
- mean-squared error (mse) in estimating the state

Optimal Transport Filter

Numerical example: Lorenz 63



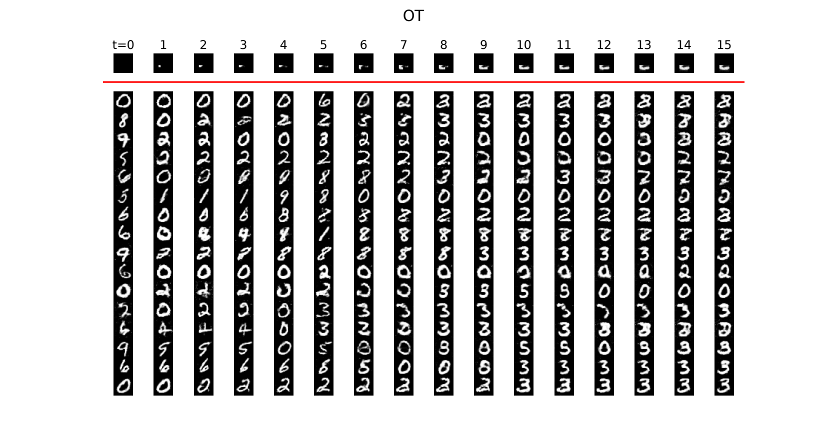
- Trajectory of the particles
- mean-squared error (mse) in estimating the state

Numerical example: Image in-painting

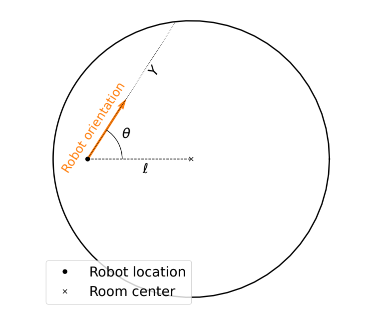
$$X \sim N(0, I_{100}),$$
$$Y_t = h(G(X), c_t) + W_t,$$
$$G : \mathbb{R}^{100} \to \mathbb{R}^{28\times 28} \text{(pre-trained generator)}$$

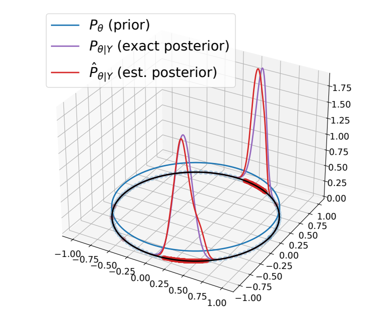
Numerical example: Image in-painting

$$X \sim N(0, I_{100}),$$
$$Y_t = h(G(X), c_t) + W_t,$$
$$G : \mathbb{R}^{100} \to \mathbb{R}^{28\times 28} \text{(pre-trained generator)}$$



Numerical example: Attitude estimation





D. Grange, M. Al-Jarrah, R. Baptista, A. Taghvaei, T. Georgiou, S. Phillips, A. Tannenbaum, Computational optimal transport and filtering on Riemannian manifolds, IEEE Control Systems Letters, 2023

Summary

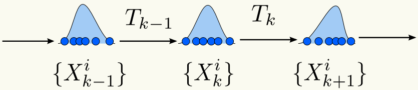
- **Mathematical model:**

$$\begin{tikzcd} {} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\ & Y_{k-1} & Y_k & Y_{k+1} & \end{tikzcd}$$

- **Nonlinear filtering:** compute the posterior $\pi_k = \mathsf{P}(X_k | Y_{1:k})$

$$\longrightarrow \pi_{k-1} \longrightarrow \pi_k \longrightarrow \pi_{k+1} \longrightarrow$$

- **OT approach:**



- **Variational problem:**

$$T_k \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N} \sum_{i=1}^{N} \delta_{(X^i_t, Y^i_t)})$$

References

- A. Taghvaei, B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms, IEEE Conference on Decision and Control (CDC), 2022
- M. Al-Jarrah, B. Hosseini, A. Taghvaei, Optimal Transport Particle Filters, IEEE Conference on Decision and Control (CDC), 2023
- M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Optimal Transport-based Nonlinear Filtering in High-dimensional Settings, arXiv:2310.13886
- B. Hosseini, AW. Hsu, A. Taghvaei, Conditional Optimal Transport on Function Spaces, arXiv:2311.05672
- D. Grange, M. Al-Jarrah, R. Baptista, A. Taghvaei, T. Georgiou, S. Phillips, A. Tannenbaum, Computational optimal transport and filtering on Riemannian manifolds, IEEE Control Systems Letters, 2023
- D. Grange, R. Baptista, A. Taghvaei, A. Tannenbaum, S. Phillips, Distributed Nonlinear Filtering using Triangular Transport Maps, IEEE American Control Conference (ACC), 2024