Optimal Transport Particle Filters

Presented at
IEEE Conference on Decision and Control, Singapore, Dec. 2023

Mohammad Al-Jarrah

Joint work with Bamdad Hosseini & Amirhossein Taghvaei

University of Washington, Seattle

Dec 15, 2023

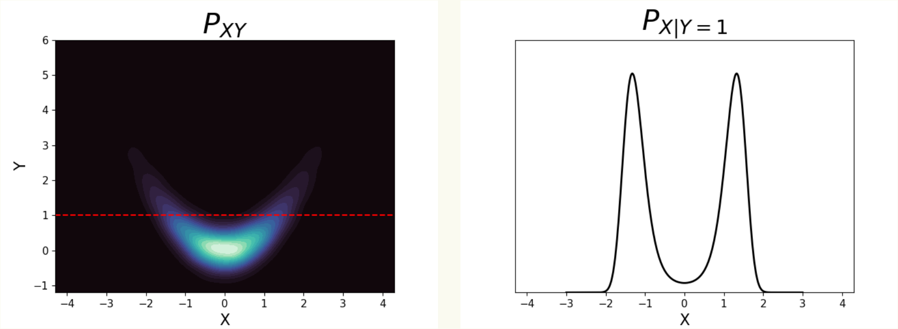
Illustrative example

- Hidden state: $X \sim N(0,1)$
- Observation: $Y = 0.5X^2 + \sigma_w W, \quad W \sim N(0,1)$

Objective: find $P(X|Y=1)$?

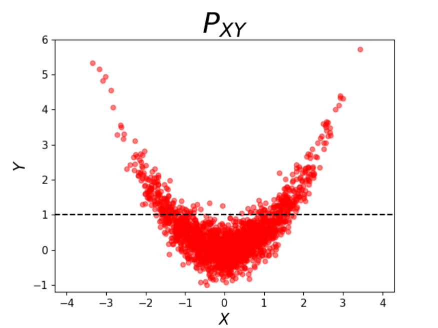
Illustrative example

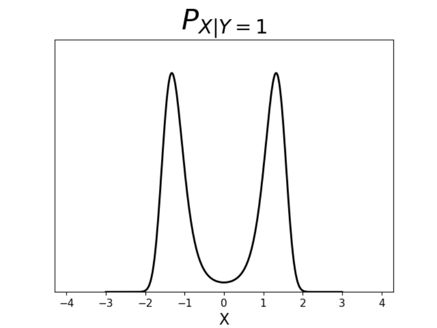
- Hidden state: $X \sim N(0, 1)$
- Observation: $Y = 0.5X^2 + \sigma_w W, \quad W \sim N(0, 1)$



Illustrative example

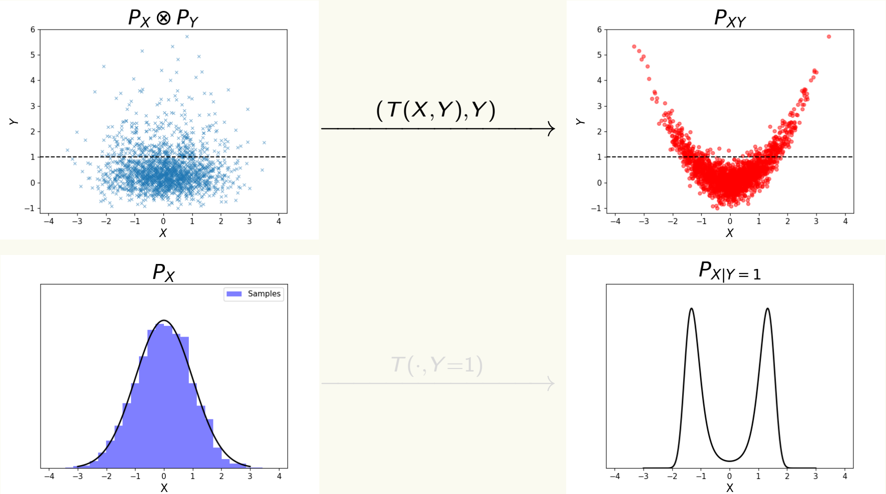
- Objective: find $P(X|Y = 1)$?
- Samples $(X_i, Y_i) \sim P_{XY}$





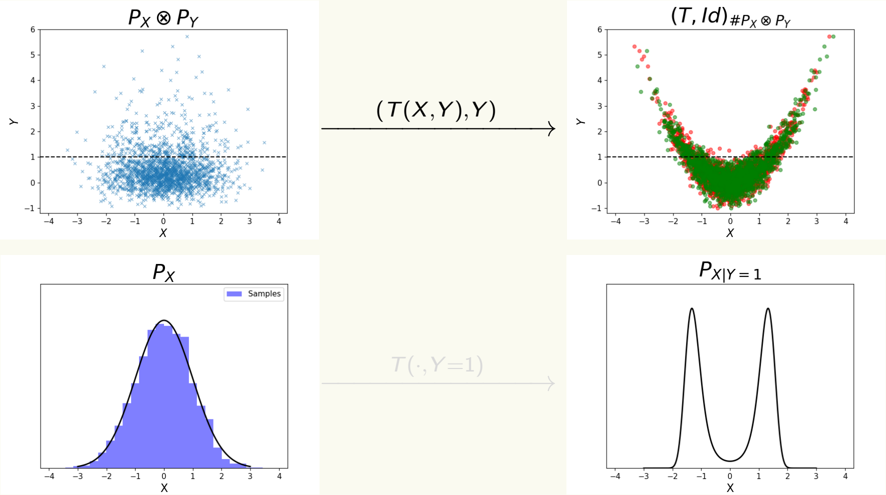
Illustrative example

- Objective: find $P(X|Y=1)$?
- Find a map $T(X,Y)$ to transport $P_X \otimes P_Y$ to P_{XY}



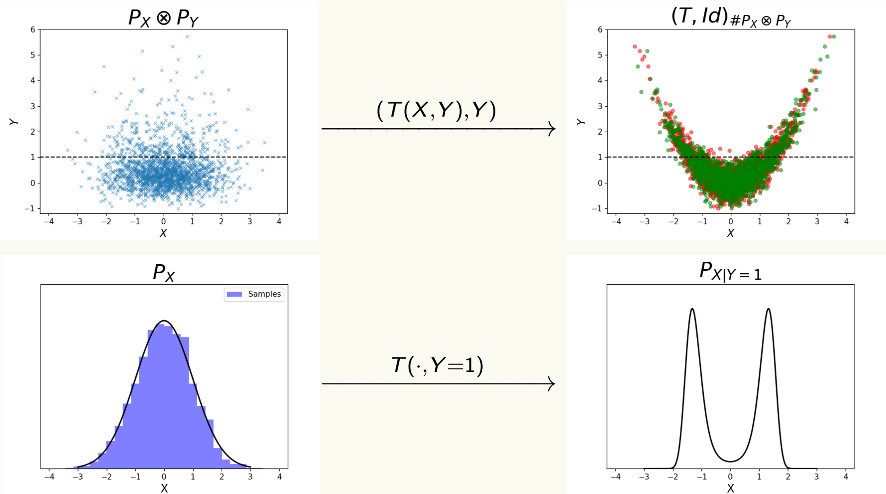
Illustrative example

- Objective: find $P(X|Y=1)$?
- Find a map $T(X,Y)$ to transport $P_X \otimes P_Y$ to P_{XY}



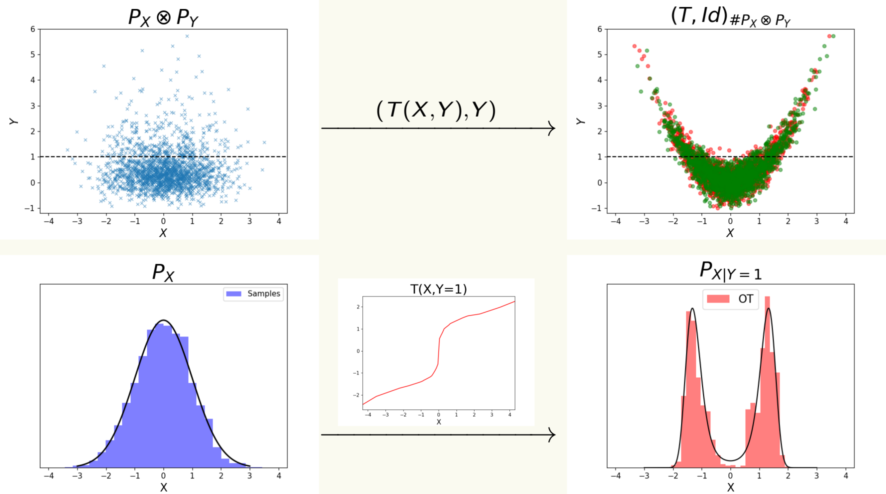
Illustrative example

- Objective: find $P(X|Y=1)$?
- Fix $Y=1$ in $T(X, Y=1)$ to transport P_X to $P_{X|Y=1}$



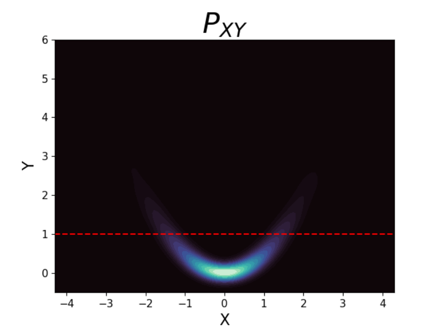
Illustrative example

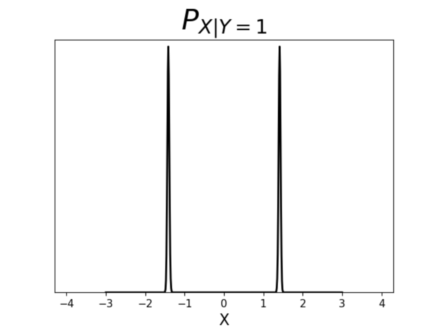
- Objective: find $P(X|Y=1)$?



Illustrative example
Degenerate likelihood

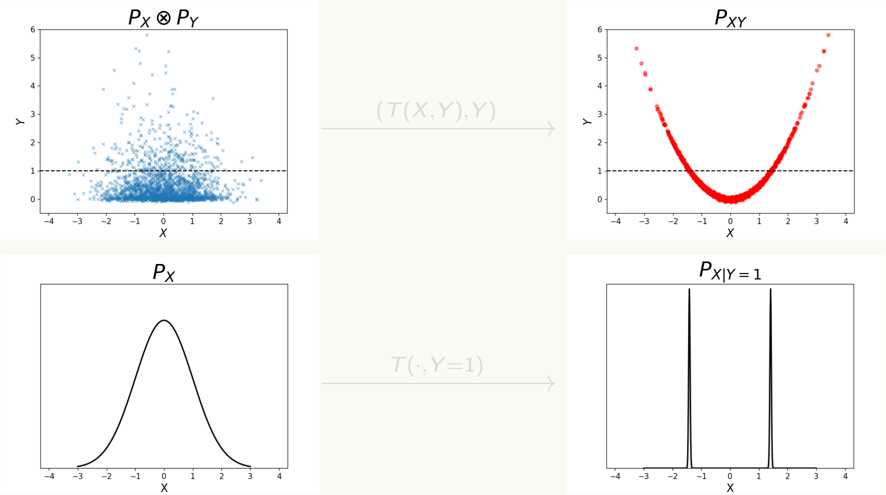
- Hidden state: $X \sim N(0, I)$
- Observation: $Y = 0.5X^2 + \sigma_w W, \quad W \sim N(0, I)$



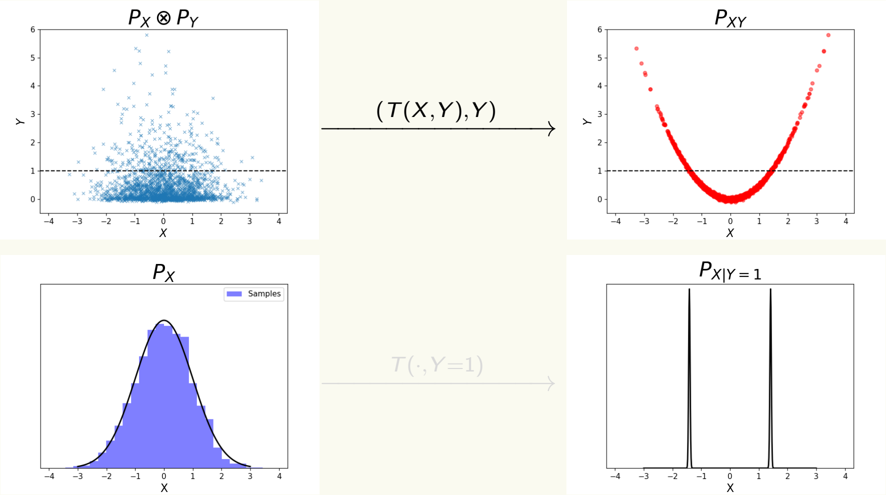


Illustrative example
Degenerate likelihood

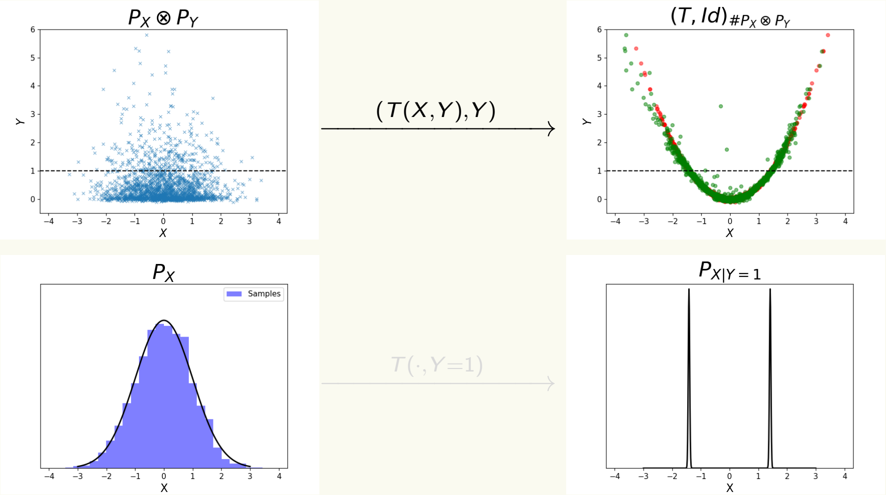
- Samples $(X_i, Y_i) \sim P_{XY}$

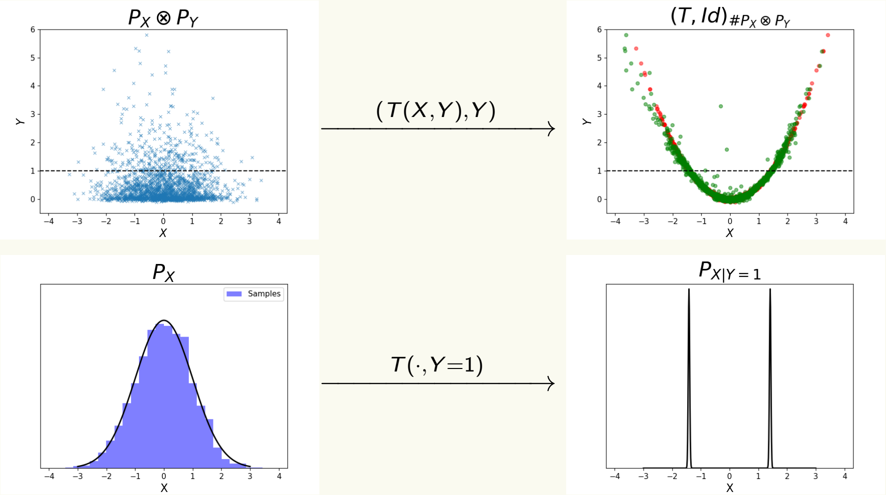


Illustrative example
Degenerate likelihood



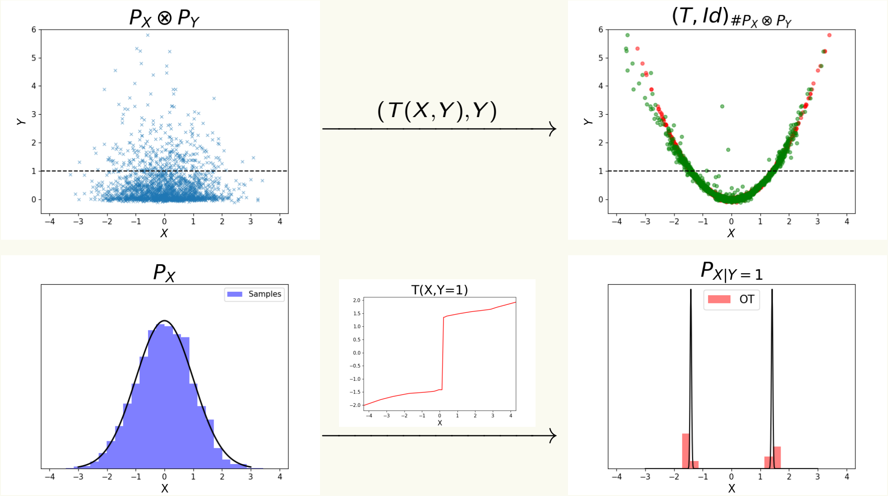
Illustrative example
Degenerate likelihood



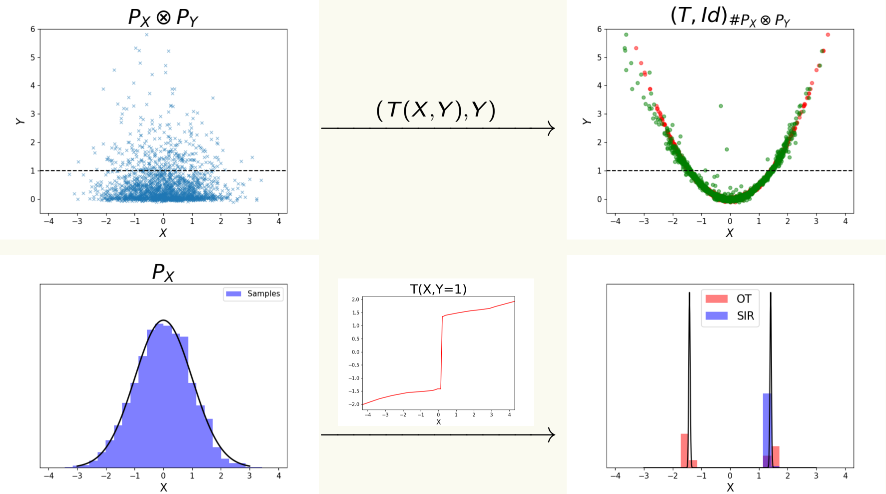


P_X ⊗ P_Y
(T, Id)_{#P_X ⊗ P_Y}
(T(X,Y),Y)
P_X
Samples
T(·,Y=1)
P_{X|Y=1}
X
Y

Illustrative example
Degenerate likelihood



Sequential Important Resampling filter suffers from weight degeneracy



Optimal transport formulation of the Bayes' law

Bayes law: $P(X|Y) = \dfrac{P(X)P(Y|X)}{P(Y)}$

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

$= \nabla_x \bar{f}(\cdot; Y)\#P_X$

where $\bar{f} = \underset{f \in L^1(\mathcal{X}\times\mathcal{Y})}{\arg\min}\ \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y}[f(X;Y)] + \mathbb{E}_{(X,Y)\sim P_{XY}}[f^\star(X;Y)]$

features:

- sample based algorithm
- stochastic optimization
- using neural network

overcome challenges:

- degenerate likelihood
- multi-model distribution
- high dimension problem

A. Taghvaei, B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms, 2022

Optimal transport formulation of the Bayes' law

Bayes law:
$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$
$$= T(\cdot; Y)\#P_X$$
$$= \nabla_x \bar{f}(\cdot; Y)\#P_X$$

where $\bar{f} = \underset{f \in L^1(\mathcal{X}\times\mathcal{Y})}{\arg\min} \ \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y}[f(X;Y)] + \mathbb{E}_{(X,Y)\sim P_{XY}}[f^\star(X;Y)]$

features:

- sample based algorithm
- stochastic optimization
- using neural network

overcome challenges:

- degenerate likelihood
- multi-model distribution
- high dimension problem

A. Taghvaei, B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms, 2022

Optimal transport formulation of the Bayes' law

Bayes law: $$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$

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

$$= \nabla_x \bar{f}(\cdot; Y)\#P_X$$

where $$\bar{f} = \underset{f \in L^1(\mathcal{X} \times \mathcal{Y})}{\arg\min}\ \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y}[f(X;Y)] + \mathbb{E}_{(X,Y)\sim P_{XY}}[f^\star(X;Y)]$$

features:

- sample based algorithm
- stochastic optimization
- using neural network

overcome challenges:

- degenerate likelihood
- multi-model distribution
- high dimension problem

A. Taghvaei, B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms, 2022

Optimal transport formulation of the Bayes' law

$$\textbf{Bayes law:} \quad P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} = T(\cdot;Y)\#P_X = \nabla_x \bar{f}(\cdot;Y)\#P_X$$

$$\text{where} \quad \bar{f} = \underset{f \in L^1(\mathcal{X}\times\mathcal{Y})}{\arg\min} \ \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y}[f(X;Y)] + \mathbb{E}_{(X,Y)\sim P_{XY}}[f^\star(X;Y)]$$

features:

- sample based algorithm
- stochastic optimization
- using neural network

overcome challenges:

- degenerate likelihood
- multi-model distribution
- high dimension problem

A. Taghvaei, B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms, 2022

Optimal transport formulation of the Bayes' law

Bayes law:
$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} = T(\cdot;Y)\#P_X = \nabla_x \bar{f}(\cdot;Y)\#P_X$$

where $$\bar{f} = \arg\min_{f \in L^1(\mathcal{X}\times\mathcal{Y})} \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y}[f(X;Y)] + \mathbb{E}_{(X,Y)\sim P_{XY}}[f^\star(X;Y)]$$

features:

- sample based algorithm
- stochastic optimization
- using neural network

overcome challenges:

- degenerate likelihood
- multi-model distribution
- high dimension problem

A. Taghvaei, B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms, 2022

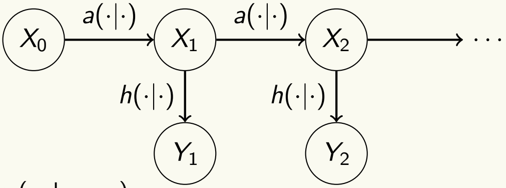
Outline

- Background on the filtering problem
- Optimal Transport Particle Filters
- Error Analysis

Outline

- Background on the filtering problem
- Optimal Transport Particle Filters
- Error Analysis

Nonlinear filtering problem
Mathematical model



- X_k is the state (unknown)
- Y_k is the observation
- have access to simulate through $a(\cdot|\cdot), h(\cdot|\cdot)$

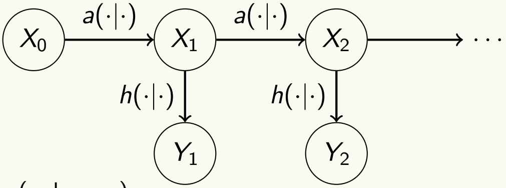
Questions: Given history of observation $Y_{1:k} := \{Y_1, \ldots, Y_k\}$,

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

Answer: given by the conditional distribution $\pi_k = \mathrm{P}(X_k | Y_{1:k})$ (posterior, belief)

J. Xiong, An introduction to stochastic filtering theory, 2008

Nonlinear filtering problem
Mathematical model



- X_k is the state (unknown)
- Y_k is the observation
- have access to simulate through $a(\cdot|\cdot), h(\cdot|\cdot)$

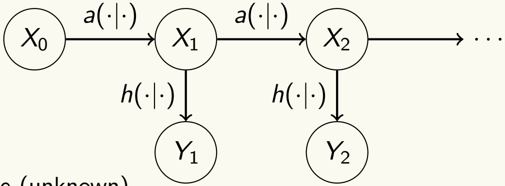
Questions: Given history of observation $Y_{1:k} := \{Y_1, \dots, Y_k\}$,

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

Answer: given by the conditional distribution $\pi_k = \mathrm{P}(X_k | Y_{1:k})$ (posterior, belief)

J. Xiong, An introduction to stochastic filtering theory, 2008

Nonlinear filtering problem
Mathematical model



- X_k is the state (unknown)
- Y_k is the observation
- have access to simulate through $a(\cdot|\cdot), h(\cdot|\cdot)$

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

- What is the most likely value of X_k? $\arg\max_x \mathsf{P}(X_k = x | Y_{1:k})$
- What is the probability of $X_k \in A$? $\int_A \mathsf{P}(X_k = x | Y_{1:k}) \mathrm{d}x$
- What is the best m.s.e estimate for X_k? $\int x \mathsf{P}(X_k = x | Y_{1:k}) \mathrm{d}x$
- ...

Answer: given by the conditional distribution $\pi_k = \mathsf{P}(X_k | Y_{1:k})$ (posterior, belief)

Recursive update for posterior

In principle: given $\pi_k = \mathrm{P}(X_k | Y_{1:k})$, obtain $\pi_{k+1} = \mathrm{P}(X_{k+1} | Y_{1:k+1})$ according to

- Step 1: propagation update

$$\pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx$$

- Step 2: conditioning update

$$\pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi_k(x)dx}$$

where $\pi_{k+1} = \mathcal{B}_{Y_k}\mathcal{A}\pi_k$

In practice: No closed-form solution except special cases (e.g. linear Gaussian)

Kalman filter fails to represent multi-modal distributions → particle filters

Particle filter exact as $N \to \infty$,but suffer from weight degeneracy in high dimension

---

R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, 1960

R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory, 1961

Recursive update for posterior

In principle: given $\pi_k = \mathrm{P}(X_k | Y_{1:k})$, obtain $\pi_{k+1} = \mathrm{P}(X_{k+1} | Y_{1:k+1})$ according to

- Step 1: propagation update

$$\pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx$$

- Step 2: conditioning update

$$\pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi_k(x)dx}$$

where $\pi_{k+1} = \mathcal{B}_{Y_k}\mathcal{A}\pi_k$

In practice: No closed-form solution except special cases (e.g. linear Gaussian)

Kalman filter fails to represent multi-modal distributions $\rightarrow$ particle filters

Particle filter exact as $N \rightarrow \infty$,but suffer from weight degeneracy in high dimension

---

R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, 1960
R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory, 1961

Recursive update for posterior

In principle: given $\pi_k = \mathrm{P}(X_k | Y_{1:k})$, obtain $\pi_{k+1} = \mathrm{P}(X_{k+1} | Y_{1:k+1})$ according to

- Step 1: propagation update

$$\pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx$$

- Step 2: conditioning update

$$\pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi_k(x)dx}$$

where $\pi_{k+1} = \mathcal{B}_{Y_k}\mathcal{A}\pi_k$

In practice: No closed-form solution except special cases (e.g. linear Gaussian)

Kalman filter fails to represent multi-modal distributions $\rightarrow$ particle filters

Particle filter exact as $N \rightarrow \infty$,but suffer from weight degeneracy in high dimension

---

R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, 1960

R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory, 1961

Recursive update for posterior

In principle: given $\pi_k = \mathrm{P}(X_k | Y_{1:k})$, obtain $\pi_{k+1} = \mathrm{P}(X_{k+1} | Y_{1:k+1})$ according to

- Step 1: propagation update

$$\pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx$$

- Step 2: conditioning update

$$\pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi_k(x)dx}$$

where $\pi_{k+1} = \mathcal{B}_{Y_k}\mathcal{A}\pi_k$

In practice: No closed-form solution except special cases (e.g. linear Gaussian)

Kalman filter fails to represent multi-modal distributions $\rightarrow$ particle filters

Particle filter exact as $N \rightarrow \infty$,but suffer from weight degeneracy in high dimension

R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, 1960
R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory, 1961

Recursive update for posterior

In principle: given $\pi_k = \mathrm{P}(X_k | Y_{1:k})$, obtain $\pi_{k+1} = \mathrm{P}(X_{k+1} | Y_{1:k+1})$ according to

- Step 1: propagation update

$$\pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx$$

- Step 2: conditioning update

$$\pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi_k(x)dx}$$

where $\pi_{k+1} = \mathcal{B}_{Y_k}\mathcal{A}\pi_k$

In practice: No closed-form solution except special cases (e.g. linear Gaussian)

Kalman filter fails to represent multi-modal distributions → particle filters

Particle filter exact as $N \to \infty$,but suffer from weight degeneracy in high dimension

R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, 1960
R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory, 1961

Recursive update for posterior

In principle: given $\pi_k = \mathsf{P}(X_k | Y_{1:k})$, obtain $\pi_{k+1} = \mathsf{P}(X_{k+1} | Y_{1:k+1})$ according to

- Step 1: propagation update

$$\pi \xrightarrow{\text{dynamics}} \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx$$

- Step 2: conditioning update

$$\pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi_k(x)dx}$$

where $\pi_{k+1} = \mathcal{B}_{Y_k}\mathcal{A}\pi_k$

In practice: No closed-form solution except special cases (e.g. linear Gaussian)

Kalman filter fails to represent multi-modal distributions $\rightarrow$ particle filters

Particle filter exact as $N \rightarrow \infty$,but suffer from weight degeneracy in high dimension

---

R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, 1960
R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory, 1961

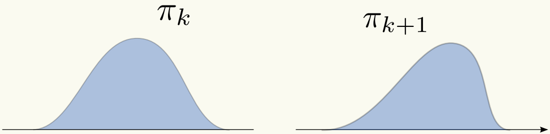
Outline

- Background on the filtering problem
- Optimal Transport Particle Filters
- Error Analysis

Outline

- Background on the filtering problem
- Optimal Transport Particle Filters
- Error Analysis

Transport view point



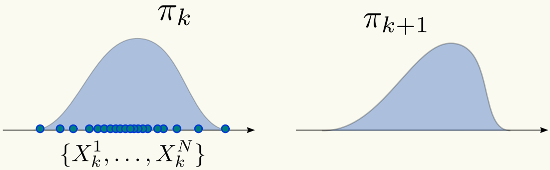
- suppose we have particles that represent samples from π_k
- we like to generate new set of particles that represent samples from π_{k+1}
- the dynamic update is straightforward, however, the Bayes update is challenging

Transport view-point: update particles with a transport map from π_k to π_{k+1}

$$X^i_{k+1} = T_k(X^i_k)$$

Question: How to numerically approximate the transport map T_k?

Transport view point



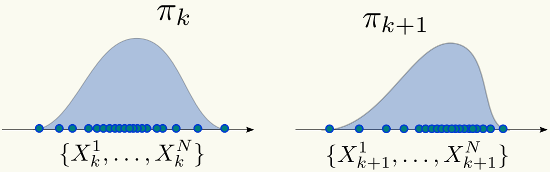
- suppose we have particles that represent samples from π_k
- we like to generate new set of particles that represent samples from π_{k+1}
- the dynamic update is straightforward, however, the Bayes update is challenging

Transport view-point: update particles with a transport map from π_k to π_{k+1}

$$X_{k+1}^i = T_k(X_k^i)$$

Question: How to numerically approximate the transport map T_k?

Transport view point



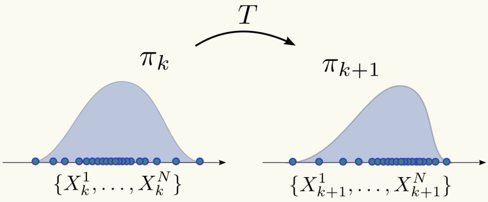
- suppose we have particles that represent samples from π_k
- we like to generate new set of particles that represent samples from π_{k+1}
- the dynamic update is straightforward, however, the Bayes update is challenging

Transport view-point: update particles with a transport map from π_k to π_{k+1}

$$X_{k+1}^i = T_k(X_k^i)$$

Question: How to numerically approximate the transport map T_k?

Transport view point



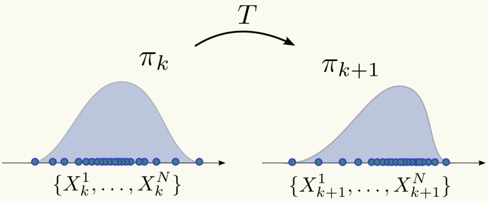
- suppose we have particles that represent samples from π_k
- we like to generate new set of particles that represent samples from π_{k+1}
- the dynamic update is straightforward, however, the Bayes update is challenging

Transport view-point: update particles with a transport map from π_k to π_{k+1}

$$X_{k+1}^i = T_k(X_k^i)$$

Question: How to numerically approximate the transport map T_k?

Transport view point



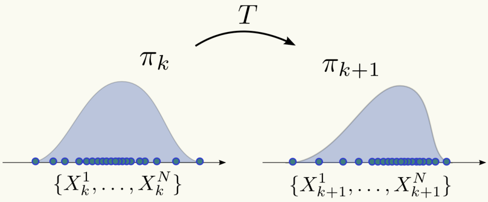
- suppose we have particles that represent samples from π_k
- we like to generate new set of particles that represent samples from π_{k+1}
- the dynamic update is straightforward, however, the Bayes update is challenging

Transport view-point: update particles with a transport map from π_k to π_{k+1}

$$X_{k+1}^i = T_k(X_k^i)$$

Question: How to numerically approximate the transport map T_k?

Transport view point



- suppose we have particles that represent samples from π_k
- we like to generate new set of particles that represent samples from π_{k+1}
- the dynamic update is straightforward, however, the Bayes update is challenging

Transport view-point: update particles with a transport map from π_k to π_{k+1}

$$X_{k+1}^i = T_k(X_k^i)$$

Question: How to numerically approximate the transport map T_k?

Optimal transport particle filter

- true observation $Y_k \sim h(\cdot|X_k)$
- given particles $\{X_k^i\}_{i=1}^N \sim \pi_k$, generate

$$Y_k^i \sim h(\cdot|X_k^i)$$

- use $\{(X_k^i, Y_k^i)\}_{i=1}^N$ to obtain $\bar{f}$ by solving

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N f(X_k^i; Y_k^{\sigma_i}) + \frac{1}{N} \sum_{i=1}^N f^\star(X_k^i; Y_k^i)$$

- where $\mathcal{F}$ is a paramteric class of functions
  - class of quadratic functions $\rightarrow$ Optimal Transport EnKF
  - subset of convex functions (e.g. ICNNs)
- update the particles according to

$$X_{k+1}^i = \nabla \bar{f}_k(X_k^i, Y_k) = T(X_k^i, Y_k)$$

Optimal transport particle filter

- true observation $Y_k \sim h(\cdot|X_k)$
- given particles $\{X_k^i\}_{i=1}^N \sim \pi_k$, generate

$$Y_k^i \sim h(\cdot|X_k^i)$$

- use $\{(X_k^i, Y_k^i)\}_{i=1}^N$ to obtain $\bar{f}$ by solving

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N f(X_k^i; Y_k^{\sigma_i}) + \frac{1}{N} \sum_{i=1}^N f^\star(X_k^i; Y_k^i)$$

- where $\mathcal{F}$ is a paramteric class of functions
  - class of quadratic functions $\rightarrow$ Optimal Transport EnKF
  - subset of convex functions (e.g. ICNNs)
- update the particles according to

$$X_{k+1}^i = \nabla \bar{f}_k(X_k^i, Y_k) = T(X_k^i, Y_k)$$

Optimal transport particle filter

- true observation $Y_k \sim h(\cdot|X_k)$
- given particles $\{X_k^i\}_{i=1}^N \sim \pi_k$, generate

$$Y_k^i \sim h(\cdot|X_k^i)$$

- use $\{(X_k^i, Y_k^i)\}_{i=1}^N$ to obtain $\bar{f}$ by solving

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N f(X_k^i; Y_k^{\sigma_i}) + \frac{1}{N} \sum_{i=1}^N f^\star(X_k^i; Y_k^i)$$

- where $\mathcal{F}$ is a paramteric class of functions
  - class of quadratic functions $\rightarrow$ Optimal Transport EnKF
  - subset of convex functions (e.g. ICNNs)
- update the particles according to

$$X_{k+1}^i = \nabla \bar{f}_k(X_k^i, Y_k) = T(X_k^i, Y_k)$$

Optimal transport particle filter

- true observation $Y_k \sim h(\cdot|X_k)$
- given particles $\{X_k^i\}_{i=1}^N \sim \pi_k$, generate

$$Y_k^i \sim h(\cdot|X_k^i)$$

- use $\{(X_k^i, Y_k^i)\}_{i=1}^N$ to obtain $\bar{f}$ by solving

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N f(X_k^i; Y_k^{\sigma_i}) + \frac{1}{N} \sum_{i=1}^N f^\star(X_k^i; Y_k^i)$$

- where $\mathcal{F}$ is a paramteric class of functions
  - class of quadratic functions $\rightarrow$ Optimal Transport EnKF
  - subset of convex functions (e.g. ICNNs)
- update the particles according to

$$X_{k+1}^i = \nabla \bar{f}_k(X_k^i, Y_k) = T(X_k^i, Y_k)$$

Optimal transport particle filter

- true observation $Y_k \sim h(\cdot|X_k)$
- given particles $\{X_k^i\}_{i=1}^N \sim \pi_k$, generate

$$Y_k^i \sim h(\cdot|X_k^i)$$

- use $\{(X_k^i, Y_k^i)\}_{i=1}^N$ to obtain $\bar{f}$ by solving

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N f(X_k^i; Y_k^{\sigma_i}) + \frac{1}{N} \sum_{i=1}^N f^\star(X_k^i; Y_k^i)$$

- where $\mathcal{F}$ is a paramteric class of functions
  - class of quadratic functions $\rightarrow$ Optimal Transport EnKF
  - subset of convex functions (e.g. ICNNs)
- update the particles according to

$$X_{k+1}^i = \nabla \bar{f}_k(X_k^i, Y_k) = T(X_k^i, Y_k)$$

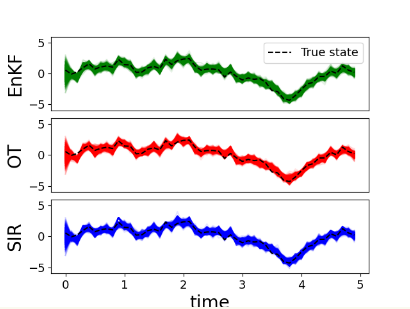
Numerical example: Dynamical model

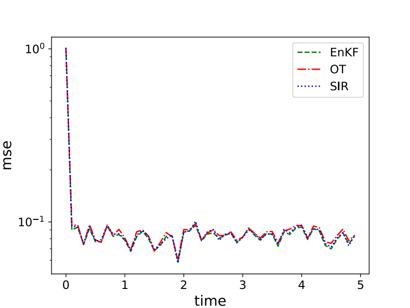
$$
\begin{aligned}
X_t &= (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),\\
Y_t &= h(X_t) + \sigma_W W_t,
\end{aligned}
$$

Numerical example: Dynamical model

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

$$h(X_t) = X_t$$

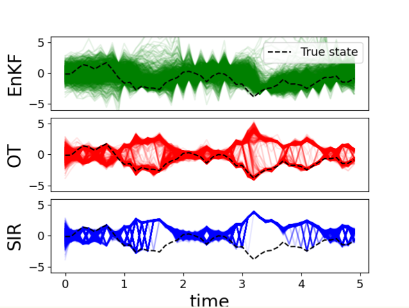


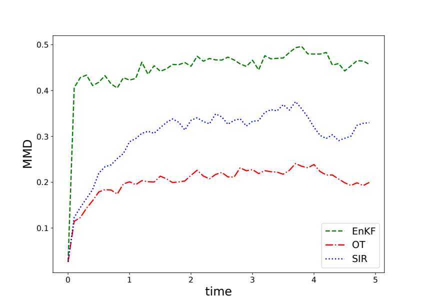


Numerical example: Dynamical model

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

$$h(X_t) = X_t^2$$





Outline

- Background on the filtering problem
- Optimal Transport Particle Filters
- Error Analysis

Outline

- Background on the filtering problem
- Optimal Transport Particle Filters
- Error Analysis

Definitions

Optimization problem:

$$\min_{f\in CVX_x} J(f,\pi) := \mathbb{E}[f(\bar{X},Y)+f^\star(X,Y)]$$

The exact process:

$$\bar{\pi}_t = \nabla_x \bar{f}_t(\cdot,Y_t)\#\mathcal{A}\bar{\pi}_{t-1} = \mathcal{B}_y\mathcal{A}\bar{\pi}_{t-1}$$
$$\bar{f}_t = \arg\min_{f\in\mathrm{CVX}_x} J(f,\mathcal{A}\bar{\pi}_{t-1})$$

The approximate mean-field process: $\mathcal{F}\subset CVX_x$

$$\pi_t^{\mathcal{F}} = \nabla_x f_t^{\mathcal{F}}(\cdot,Y_t)\#\mathcal{A}\pi_{t-1}^{\mathcal{F}}$$
$$f_t^{\mathcal{F}} = \arg\min_{f\in\mathcal{F}} J(f,\mathcal{A}\pi_{t-1}^{\mathcal{F}})$$

The finite particle system: $\mathcal{S}$ is a sampling operator

$$\tilde{\pi}_t^{(\mathcal{F},N)} = \nabla_x \tilde{f}_t^{(\mathcal{F},N)}(\cdot,Y_t)\#\mathcal{S}^N\mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)}$$
$$\tilde{f}_t^{(\mathcal{F},N)} = \arg\min_{f\in\mathcal{F}} J(f,\mathcal{S}^N\mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)})$$

Definitions

Optimization problem:

$$\min_{f \in CVX_x} J(f, \pi) := \mathbb{E}[f(\bar{X}, Y) + f^\star(X, Y)]$$

The exact process:

$$\bar{\pi}_t = \nabla_x \bar{f}_t(\cdot, Y_t) \# \mathcal{A}\bar{\pi}_{t-1} = \mathcal{B}_y \mathcal{A}\bar{\pi}_{t-1}$$
$$\bar{f}_t = \arg\min_{f \in \mathrm{CVX}_x} J(f, \mathcal{A}\bar{\pi}_{t-1})$$

The approximate mean-field process: $\mathcal{F} \subset CVX_x$

$$\pi_t^{\mathcal{F}} = \nabla_x f_t^{\mathcal{F}}(\cdot, Y_t) \# \mathcal{A}\pi_{t-1}^{\mathcal{F}}$$
$$f_t^{\mathcal{F}} = \arg\min_{f \in \mathcal{F}} J(f, \mathcal{A}\pi_{t-1}^{\mathcal{F}})$$

The finite particle system: $\mathcal{S}$ is a sampling operator

$$\tilde{\pi}_t^{(\mathcal{F},N)} = \nabla_x \tilde{f}_t^{(\mathcal{F},N)}(\cdot, Y_t) \# \mathcal{S}^N \mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)}$$
$$\tilde{f}_t^{(\mathcal{F},N)} = \arg\min_{f \in \mathcal{F}} J(f, \mathcal{S}^N \mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)})$$

Definitions

Optimization problem:

$$\min_{f \in CVX_x} J(f,\pi) := \mathbb{E}[f(\bar{X}, Y) + f^{\star}(X, Y)]$$

The exact process:

$$\bar{\pi}_t = \nabla_x \bar{f}_t(\cdot, Y_t) \# \mathcal{A}\bar{\pi}_{t-1} = \mathcal{B}_y \mathcal{A}\bar{\pi}_{t-1}$$
$$\bar{f}_t = \arg\min_{f \in \mathrm{CVX}_x} J(f, \mathcal{A}\bar{\pi}_{t-1})$$

The approximate mean-field process: $\mathcal{F} \subset CVX_x$

$$\pi_t^{\mathcal{F}} = \nabla_x f_t^{\mathcal{F}}(\cdot, Y_t) \# \mathcal{A}\pi_{t-1}^{\mathcal{F}}$$
$$f_t^{\mathcal{F}} = \arg\min_{f \in \mathcal{F}} J(f, \mathcal{A}\pi_{t-1}^{\mathcal{F}})$$

The finite particle system: $\mathcal{S}$ is a sampling operator

$$\tilde{\pi}_t^{(\mathcal{F},N)} = \nabla_x \tilde{f}_t^{(\mathcal{F},N)}(\cdot, Y_t) \# \mathcal{S}^N \mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)}$$
$$\tilde{f}_t^{(\mathcal{F},N)} = \arg\min_{f \in \mathcal{F}} J(f, \mathcal{S}^N \mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)})$$

Definitions

Optimization problem:

$$\min_{f \in CVX_x} J(f, \pi) := \mathbb{E}[f(\bar{X}, Y) + f^\star(X, Y)]$$

The exact process:

$$\bar{\pi}_t = \nabla_x \bar{f}_t(\cdot, Y_t) \# \mathcal{A}\bar{\pi}_{t-1} = \mathcal{B}_y \mathcal{A}\bar{\pi}_{t-1}$$
$$\bar{f}_t = \arg\min_{f \in \mathrm{CVX}_x} J(f, \mathcal{A}\bar{\pi}_{t-1})$$

The approximate mean-field process: $\mathcal{F} \subset CVX_x$

$$\pi_t^{\mathcal{F}} = \nabla_x f_t^{\mathcal{F}}(\cdot, Y_t) \# \mathcal{A}\pi_{t-1}^{\mathcal{F}}$$
$$f_t^{\mathcal{F}} = \arg\min_{f \in \mathcal{F}} J(f, \mathcal{A}\pi_{t-1}^{\mathcal{F}})$$

The finite particle system: $\mathcal{S}$ is a sampling operator

$$\tilde{\pi}_t^{(\mathcal{F},N)} = \nabla_x \tilde{f}_t^{(\mathcal{F},N)}(\cdot, Y_t) \# \mathcal{S}^N \mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)}$$
$$\tilde{f}_t^{(\mathcal{F},N)} = \arg\min_{f \in \mathcal{F}} J(f, \mathcal{S}^N \mathcal{A}\tilde{\pi}_{t-1}^{(\mathcal{F},N)})$$

Error Analysis

Theorem

Consider the exact distribution $\bar{\pi}_t$ and the particle distribution $\tilde{\pi}_t^{(\mathcal{F},N)}$. Assume

1. The exact filter is "uniformly geometrically stable".
2. The optimality gap between $J(f, \mathcal{S}^N \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ and $J(f, \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ is uniformly bounded by $\epsilon_{\mathcal{F},N}$ for all t and N.
3. For all y, t, and N, the function $f_t^{(\mathcal{F},N)}(\cdot, y)$ is convex and $\nabla_x f_t^{(\mathcal{F},N)}(\cdot, y)$ is β-Lipschitz.

Then, it holds that

$$d(\tilde{\pi}_t^{(\mathcal{F},N)}, \pi_t) \leq C \left(\sqrt{2\beta\epsilon_{\mathcal{F},N}} + \frac{1}{\sqrt{N}} \right), \quad \forall t,$$

where all constants are time-independent.

Error Analysis

Theorem

Consider the exact distribution $\bar{\pi}_t$ and the particle distribution $\tilde{\pi}_t^{(\mathcal{F},N)}$. Assume

1. The exact filter is "uniformly geometrically stable".
2. The optimality gap between $J(f, \mathcal{S}^N \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ and $J(f, \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ is uniformly bounded by $\epsilon_{\mathcal{F},N}$ for all t and N.
3. For all y, t, and N, the function $f_t^{(\mathcal{F},N)}(\cdot, y)$ is convex and $\nabla_x f_t^{(\mathcal{F},N)}(\cdot, y)$ is β-Lipschitz.

Then, it holds that

$$d(\tilde{\pi}_t^{(\mathcal{F},N)}, \pi_t) \leq C \left(\sqrt{2\beta\epsilon_{\mathcal{F},N}} + \frac{1}{\sqrt{N}} \right), \quad \forall t,$$

where all constants are time-independent.

Error Analysis

Theorem

Consider the exact distribution $\bar{\pi}_t$ and the particle distribution $\tilde{\pi}_t^{(\mathcal{F},N)}$. Assume

1. The exact filter is "uniformly geometrically stable".
2. The optimality gap between $J(f, \mathcal{S}^N \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ and $J(f, \mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)})$ is uniformly bounded by $\epsilon_{\mathcal{F},N}$ for all t and N.
3. For all y, t, and N, the function $f_t^{(\mathcal{F},N)}(\cdot, y)$ is convex and $\nabla_x f_t^{(\mathcal{F},N)}(\cdot, y)$ is β-Lipschitz.

Then, it holds that

$$d(\tilde{\pi}_t^{(\mathcal{F},N)}, \pi_t) \leq C\left(\sqrt{2\beta\epsilon_{\mathcal{F},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t,$$

where all constants are time-independent.

Error Analysis

Theorem

Consider the exact distribution $\bar{\pi}_t$ and the particle distribution $\tilde{\pi}_t^{(\mathcal{F},N)}$. Assume

1. The exact filter is "uniformly geometrically stable".
2. The optimality gap between $J(f, \mathcal{S}^N \mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)})$ and $J(f, \mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)})$ is uniformly bounded by $\epsilon_{\mathcal{F},N}$ for all t and N.
3. For all y, t, and N, the function $f_t^{(\mathcal{F},N)}(\cdot, y)$ is convex and $\nabla_x f_t^{(\mathcal{F},N)}(\cdot, y)$ is β-Lipschitz.

Then, it holds that

$$d(\tilde{\pi}_t^{(\mathcal{F},N)}, \pi_t) \leq C\left(\sqrt{2\beta\epsilon_{\mathcal{F},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t,$$

where all constants are time-independent.

Error Analysis

Theorem

Consider the exact distribution $\bar{\pi}_t$ and the particle distribution $\tilde{\pi}_t^{(\mathcal{F},N)}$. Assume

1. The exact filter is "uniformly geometrically stable".
2. The optimality gap between $J(f, \mathcal{S}^N \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ and $J(f, \mathcal{A} \tilde{\pi}_t^{(\mathcal{F},N)})$ is uniformly bounded by $\epsilon_{\mathcal{F},N}$ for all t and N.
3. For all y, t, and N, the function $f_t^{(\mathcal{F},N)}(\cdot, y)$ is convex and $\nabla_x f_t^{(\mathcal{F},N)}(\cdot, y)$ is β-Lipschitz.

Then, it holds that

$$d(\tilde{\pi}_t^{(\mathcal{F},N)}, \pi_t) \leq C \left(\sqrt{2\beta\epsilon_{\mathcal{F},N}} + \frac{1}{\sqrt{N}} \right), \quad \forall t,$$

where all constants are time-independent.

Summary

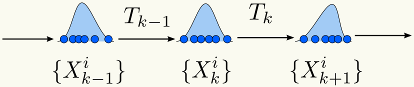
- **Mathematical model:**

$$\longrightarrow X_{k-1} \longrightarrow X_k \longrightarrow X_{k+1} \longrightarrow$$
$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$
$$Y_{k-1} \qquad Y_k \qquad Y_{k+1}$$

- **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:**

$$T_k = \nabla_x \bar{f}_k, \quad \text{where} \quad \bar{f}_k = \arg\min_{f \in \mathcal{F}} J^{(N)}(f; \{(X^i_k, Y^i_k)\})$$

Optimal transportation methods in nonlinear filtering: The feedback particle filter, IEEE CSM, 2021

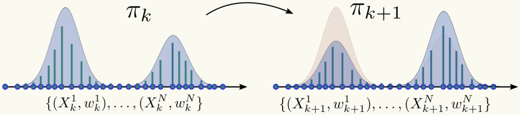
THANK YOU

ANY QUESTIONS?

Particle filters

Monte-Carlo approximation

- approximate π_k with weighted empirical distribution of particles
- apply the update rule to the particles and weights



- Step 1: update the weights according to Bayes rule

$$w_{k+1}^i \propto w_k^i h(Y_k | X_k^i)$$

- Step 2: update particles according to the dynamics

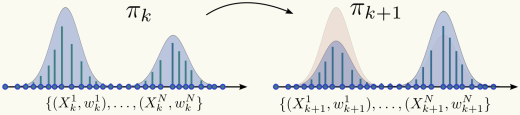
Properties:

- exact in the limit as $N \to \infty$
- weight degeneracy $\to$ curse of dimensionality

N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

Particle filters
Monte-Carlo approximation

- approximate π_k with weighted empirical distribution of particles
- apply the update rule to the particles and weights



- Step 1: update the weights according to Bayes rule

$$w_{k+1}^i \propto w_k^i h(Y_k | X_k^i)$$

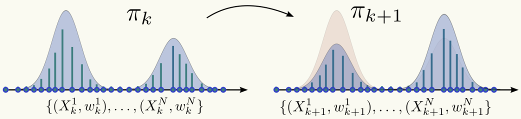
- Step 2: update particles according to the dynamics

Properties:

- exact in the limit as $N \to \infty$
- weight degeneracy $\to$ curse of dimensionality

N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

- approximate π_k with weighted empirical distribution of particles
- apply the update rule to the particles and weights



- Step 1: update the weights according to Bayes rule

$$w_{k+1}^i \propto w_k^i h(Y_k | X_k^i)$$

- Step 2: update particles according to the dynamics

Properties:

- exact in the limit as $N \to \infty$
- weight degeneracy $\to$ curse of dimensionality

N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

Derivation of the variational formula

- We want to find a map T that transports P_X to $P_{X|Y}$ with minimum cost

$$\min_{T} \mathbb{E}_{X\sim P_X}[\|T(X)-X\|^2], \quad \text{s.t.} \quad T\#P_X = P_{X|Y}$$

- The Kantorovich dual formulation removes the constraint

$$\min_{f\in L^1(\mathcal{X})} \mathbb{E}_{X\sim P_X}[f(X)] + \mathbb{E}_{X\sim P_{X|Y}}[f^\star(X)] \quad \text{but } P_{X|Y} \text{ is not available}$$

- Take expectation with respect to Y

$$\min_{f\in L^1(\mathcal{X}\times\mathcal{Y})} \mathbb{E}_{(X,Y)\sim P_X\otimes P_Y}[f(X;Y)] + \mathbb{E}_{(X,Y)\sim P_{XY}}[f^\star(X;Y)]$$

Theorem

Assume $\mathbb{E}[\|X\|^2] < \infty$ and P_X admits density.
Then, the variational problem admits a unique solution $\bar{f}$ that satisfies:

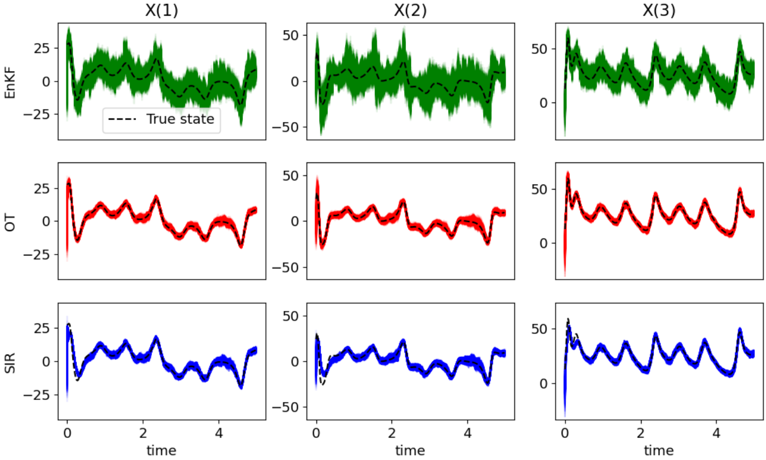
$$P_{X|Y} = \nabla_x \bar{f}(\cdot;Y)\#P_X, \quad \text{(a.e.)}$$

Numerical example: Lorenz 63 model

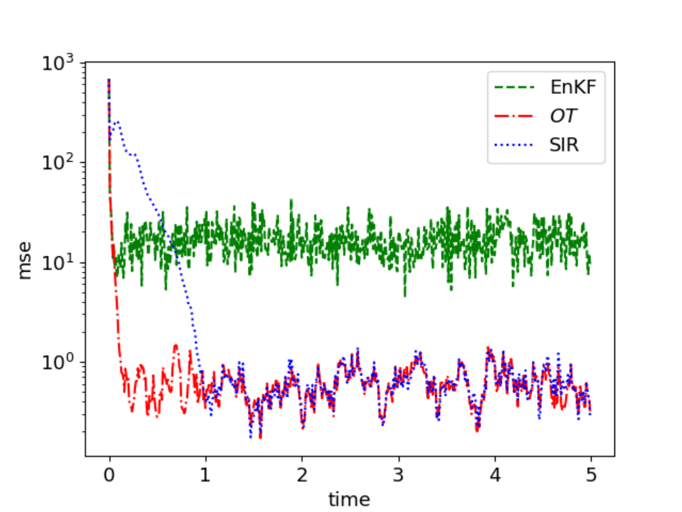
$$\dot{X} = f(X), \quad X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2 I_3),$$

$$Y_t = \begin{bmatrix} X_t(1) \\ X_t(3) \end{bmatrix} + W_t, \quad W_t \sim \mathcal{N}(0, \sigma I_2)$$

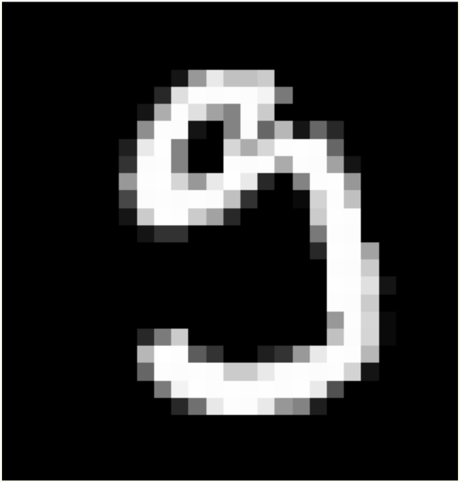
Numerical example: Lorenz 63 model



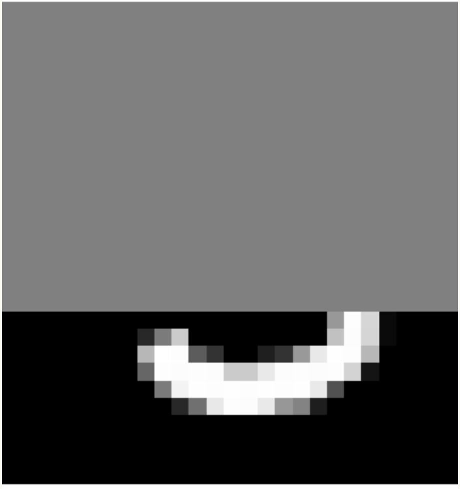
Numerical example: Lorenz 63 model



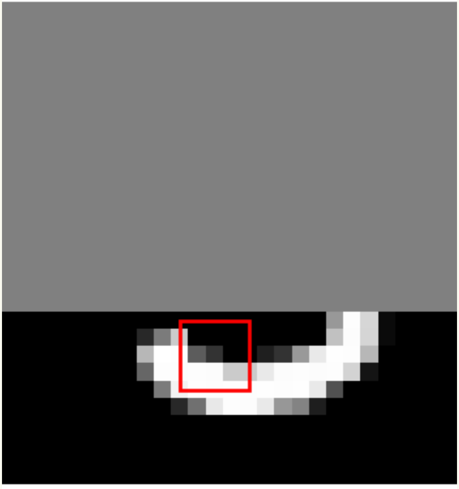
Numerical example: MNIST Dataset



Numerical example: MNIST Dataset



Numerical example: MNIST Dataset



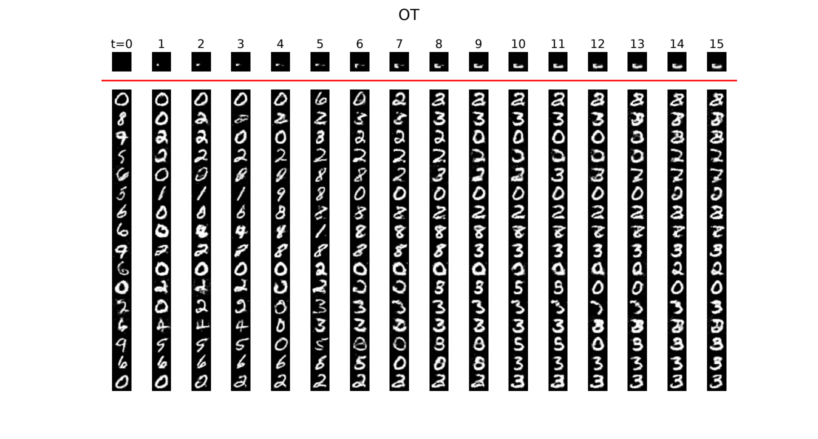
Numerical example: MNIST Dataset

$$G : \mathbb{R}^{100} \to \mathbb{R}^{28\times 28}, \quad X \sim N(0, I_{100})$$
$$Y_t = h(G(X), c_t) + W_t, \quad W_t \sim N(0, \sigma^2 I_{r^2})$$

Numerical example: MNIST Dataset

$$G : \mathbb{R}^{100} \to \mathbb{R}^{28\times 28}, \quad X \sim N(0, I_{100})$$

$$Y_t = h(G(X), c_t) + W_t, \quad W_t \sim N(0, \sigma^2 I_{r^2})$$



Numerical example: Dynamic example on MNIST Dataset

- Model:

$$X_{t+1} = (1-\alpha)X_t + V_t, \quad V_t \sim N(0, \sigma_V^2 I_{100})$$

$$Y_{t+1} = h(G(X_{t+1}), c_{t+1}) + W_{t+1}, \quad W_t \sim N(0, \sigma_W^2 I_{r^2})$$