Nonlinear Filtering with Brenier Optimal Transport Maps

Presented at
Forty-first International Conference on Machine Learning (ICML)
at Messe Wien Exhibition Congress Center, Vienna, Austria, July 2024

Mohammad Al-Jarrah

Joint work with Niyizhen "Jenny" Jin, Bamdad Hosseini, and Amirhossein Taghvaei

University of Washington, Seattle

July 2024

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

Bayes' law

Problem:

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

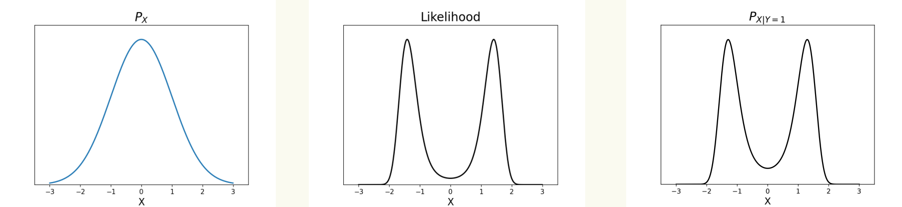
Bayes' law: $$P_{X|Y} = \frac{P_X P_{Y|X}}{P_Y}$$

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

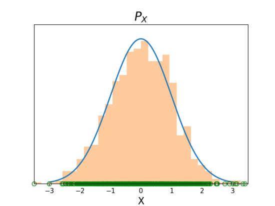
Bayes' law
Challenges of importance sampling

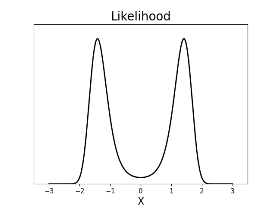
Example:

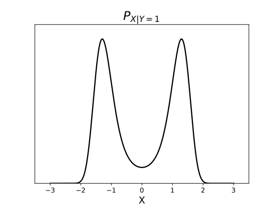
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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

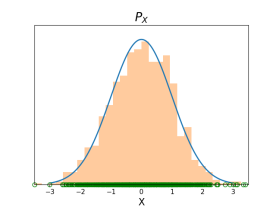
Bayes' law
Challenges of importance sampling

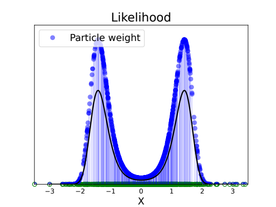
Example:

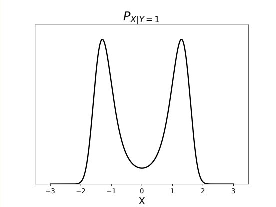
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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: $\epsilon \to 0$

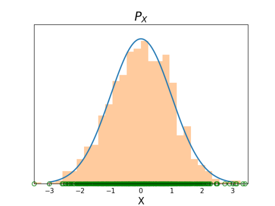
This is the main reason for the curse of dimensionality of IS-based particle filters

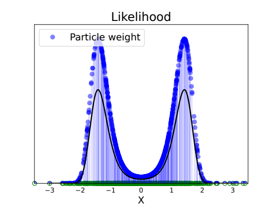
Example:

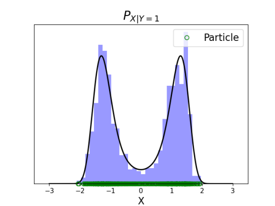
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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: $\epsilon \to 0$

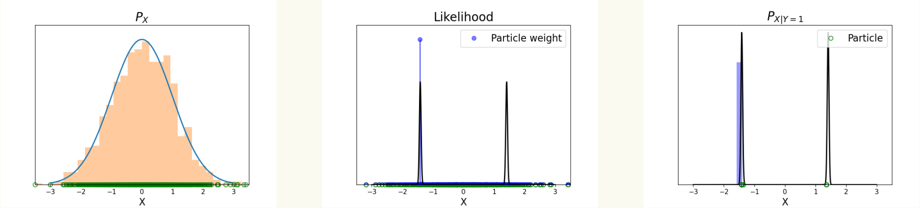
This is the main reason for the curse of dimensionality of IS-based particle filters

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: $\epsilon \to 0$

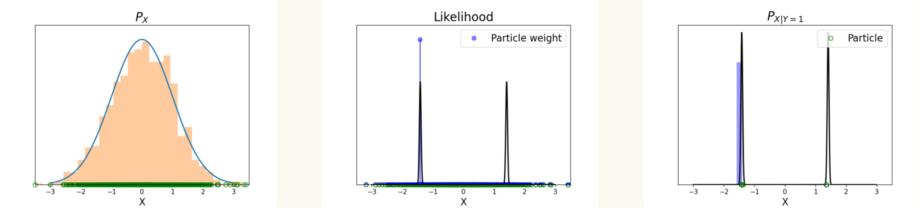
This is the main reason for the curse of dimensionality of IS-based particle filters

Example:

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

Importance sampling (IS):

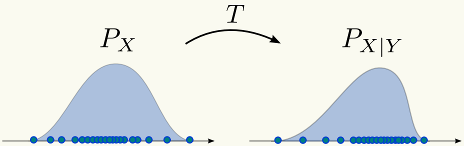
- $X^i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$
- $w^i \propto P(Y=1|X^i)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$



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

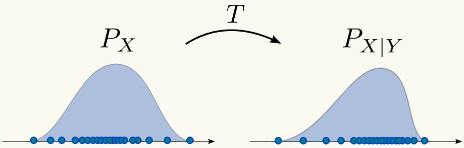


$$X^i \sim P_X \ \longrightarrow \ T(X^i, y) \sim P_{X|Y=y}$$

- Suppose we have particles that represent samples from P_X
- We like to generate new set of particles that represent samples from $P_{X|Y}$

How to numerically find the map T in a general setting?

Conditioning with transport maps

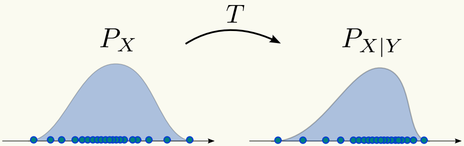


$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

- Suppose we have particles that represent samples from P_X
- We like to generate new set of particles that represent samples from $P_{X|Y}$

How to numerically find the map T in a general setting?

Conditioning with transport maps

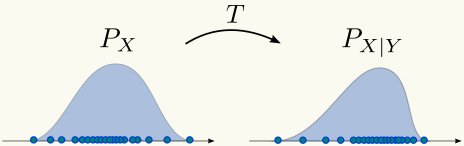


$$X^i \sim P_X \ \longrightarrow \ T(X^i, y) \sim P_{X|Y=y}$$

- Suppose we have particles that represent samples from P_X
- We like to generate new set of particles that represent samples from $P_{X|Y}$

How to numerically find the map T in a general setting?

Conditioning with transport maps



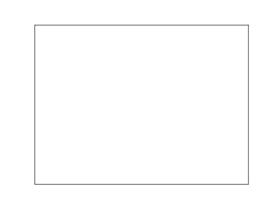
$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

- Suppose we have particles that represent samples from P_X
- We like to generate new set of particles that represent samples from $P_{X|Y}$

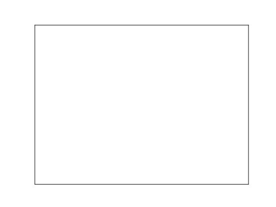
How to numerically find the map T in a general setting?

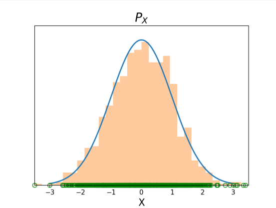
Conditioning with optimal transport map

Illustrative example

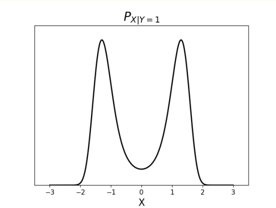


$\rightarrow$

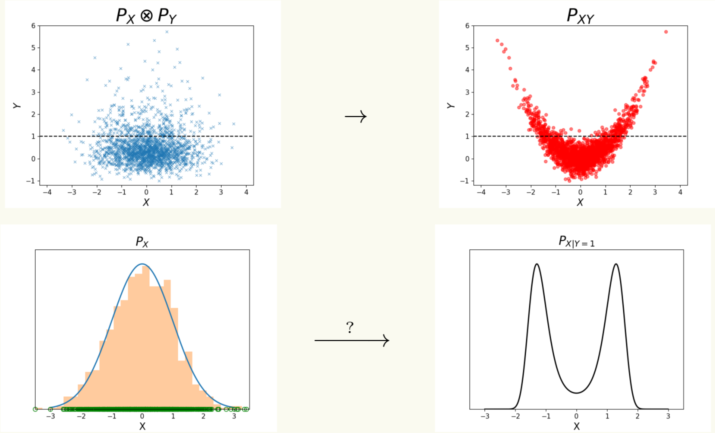




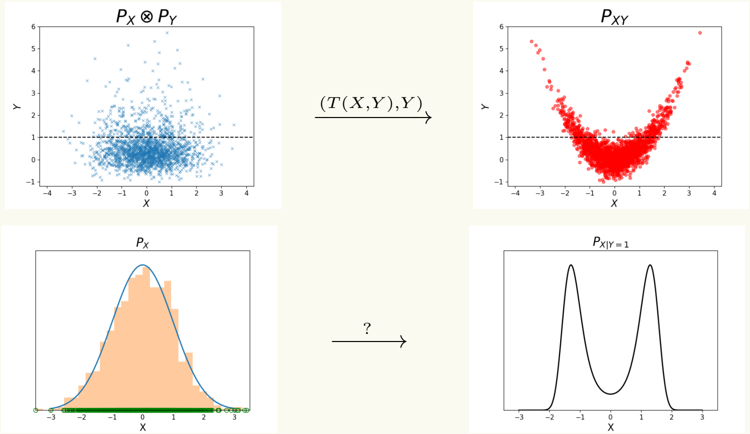
$\xrightarrow{?}$

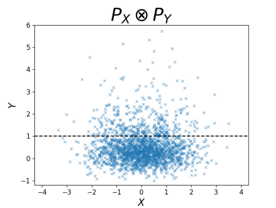


Conditioning with optimal transport map
Illustrative example

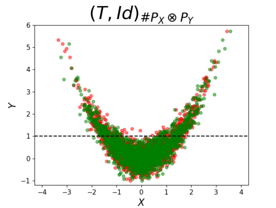


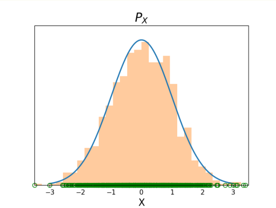
Conditioning with optimal transport map
Illustrative example



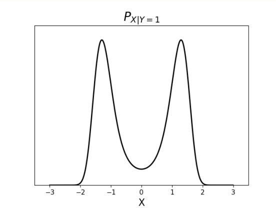


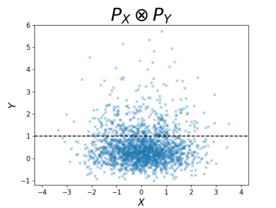
$$\xrightarrow{(T(X,Y),Y)}$$



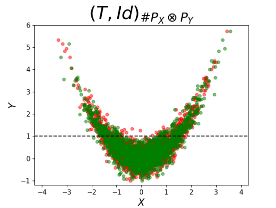


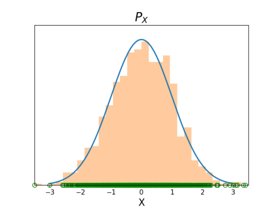
$$\xrightarrow{?}$$

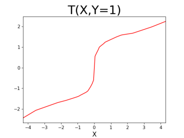




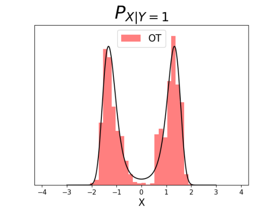
$$\xrightarrow{(T(X,Y),Y)}$$





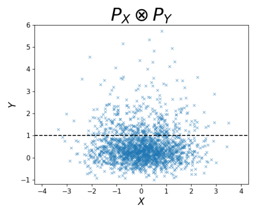


$$\longrightarrow$$

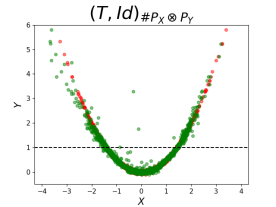


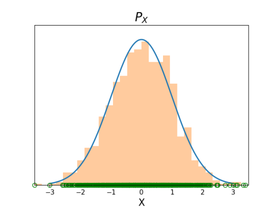
Conditioning with optimal transport map

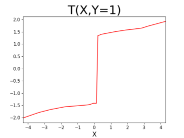
Illustrative example

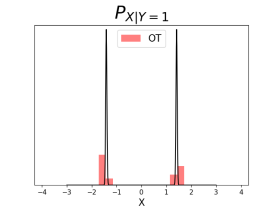


$\xrightarrow{(T(X,Y),Y)}$









small noise limit

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

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

Features:

- sample-based algorithm
- stochastic optimization
- using neural network

Overcome challenges:

- degenerate likelihood
- Multimodal distribution
- high dimension problem

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

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

Features:

- sample-based algorithm
- stochastic optimization
- using neural network

Overcome challenges:

- degenerate likelihood
- Multimodal distribution
- high dimension problem

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

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

Features:

- sample-based algorithm
- stochastic optimization
- using neural network

Overcome challenges:

- degenerate likelihood
- Multimodal distribution
- high dimension problem

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

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

Features:

- sample-based algorithm
- stochastic optimization
- using neural network

Overcome challenges:

- degenerate likelihood
- Multimodal distribution
- high dimension problem

Conditioning with optimal transport map

Variational 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

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

Features:

- sample-based algorithm
- stochastic optimization
- using neural network

Overcome challenges:

- degenerate likelihood
- Multimodal distribution
- high dimension problem

Poster session details

Our session: Wed, Jul 24, 1:30 p.m. CEST — 3 p.m. CEST
Poster Session 4
Hall C 4-9 #1409

ICML
International Conference
On Machine Learning