Data-Driven Nonlinear Filtering Algorithms with Optimal Transport Maps

Presented at the SIAM Conference on Computational Science and Engineering (CSE25)

Amirhossein Taghvaei

Joint work with Mohammad Al-Jarrah, Niyizhen Jin, and Bamdad Hosseini

Department of Aeronautics & Astronautics
University of Washington, Seattle

March 4, 2025

This talk

References:

- *Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps*
  Mohammad Al-Jarrah, Bamdad Hosseini, Amirhossein Taghvaei
  IEEE Conference on Decision and Control (CDC), Milan, 2024
- *Nonlinear Filtering with Brenier Optimal Transport Maps*
  Mohammad Al-Jarrah, Niyizhen Jin, Bamdad Hosseini, Amirhossein Taghvaei
  International Conference of Machine Learning (ICML), Vienna, 2024
- Conditional Optimal Transport on Function Spaces
  Bamdad Hosseini, Alexander W. Hsu, Amirhossein Taghvaei
  SIAM/ASA Journal on Uncertainty Quantification (Accepted)
- *Optimal Transport Particle Filters*
  Mohammad Al-Jarrah, Amirhossein Taghvaei, Bamdad Hosseini
  IEEE Conference on Decision and Control (CDC), Singapore, 2023
- *An optimal transport formulation of Bayes' law for nonlinear filtering algorithms*
  Amirhossein Taghvaei, Bamdad Hosseini
  IEEE Conference on Decision and Control (CDC), Cancun, 2022

nonlinear filtering $\xrightarrow{\text{Optimal Transport}}$ machine learning

This talk

References:

- *Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps*
  Mohammad Al-Jarrah, Bamdad Hosseini, Amirhossein Taghvaei
  IEEE Conference on Decision and Control (CDC), Milan, 2024
- *Nonlinear Filtering with Brenier Optimal Transport Maps*
  Mohammad Al-Jarrah, Niyizhen Jin, Bamdad Hosseini, Amirhossein Taghvaei
  International Conference of Machine Learning (ICML), Vienna, 2024
- Conditional Optimal Transport on Function Spaces
  Bamdad Hosseini, Alexander W. Hsu, Amirhossein Taghvaei
  SIAM/ASA Journal on Uncertainty Quantification (Accepted)
- *Optimal Transport Particle Filters*
  Mohammad Al-Jarrah, Amirhossein Taghvaei, Bamdad Hosseini
  IEEE Conference on Decision and Control (CDC), Singapore, 2023
- *An optimal transport formulation of Bayes' law for nonlinear filtering algorithms*
  Amirhossein Taghvaei, Bamdad Hosseini
  IEEE Conference on Decision and Control (CDC), Cancun, 2022

$$\text{nonlinear filtering} \xrightarrow{\text{Optimal Transport}} \text{machine learning}$$

Outline

- **Part I:** Nonlinear filtering problem
- **Part II:** Optimal transport filter
- **Part III:** Extension to data-driven setting

Outline

- **Part I:** Nonlinear filtering problem
- **Part II:** Optimal transport filter
- **Part III:** Extension to data-driven setting

Nonlinear filtering problem

Model:

$$X_t \sim a(\cdot \mid X_{t-1}), \quad X_0 \sim \pi_0$$
$$Y_t \sim h(\cdot \mid X_t)$$

- X_t is the state
- Y_t is the observation
- dynamic and observation models are available as simulators

Nonlinear filtering: numerical approximation of the posterior $\pi_t = P_{X_t|Y_1,\ldots,Y_t}$ for all t.

Nonlinear filtering problem

Model:

$$X_t \sim a(\cdot \mid X_{t-1}), \quad X_0 \sim \pi_0$$
$$Y_t \sim h(\cdot \mid X_t)$$

- X_t is the state
- Y_t is the observation
- dynamic and observation models are available as simulators

Nonlinear filtering: numerical approximation of the posterior $\pi_t = P_{X_t|Y_1,\ldots,Y_t}$ for all t.

Filtering equations

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

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

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \pi_{t|t-1} := \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} \pi_t = \mathcal{B}_{Y_t}(\pi_{t|t-1})$$

Challenge: numerical implementation of the Bayes' law

Filtering equations

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

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

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \pi_{t|t-1} := \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} \pi_t = \mathcal{B}_{Y_t}(\pi_{t|t-1})$$

Challenge: numerical implementation of the Bayes' law

Filtering equations

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

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

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y(\pi)$$

- Recursive update law for the posterior

$$\pi_{t-1} \xrightarrow{\text{dynamics}} \pi_{t|t-1} := \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} \pi_t = \mathcal{B}_{Y_t}(\pi_{t|t-1})$$

Challenge: numerical implementation of the Bayes' law

Filtering equations

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

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

$$\text{Conditioning:} \quad \pi \xrightarrow{\text{Bayes law}} \mathcal{B}_y(\pi)$$

- Recursive update law for the posterior

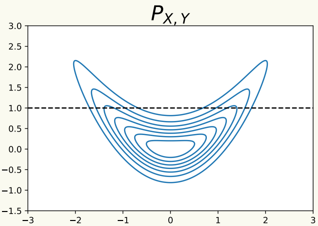
$$\pi_{t-1} \xrightarrow{\text{dynamics}} \pi_{t|t-1} := \mathcal{A}\pi_{t-1} \xrightarrow{\text{Bayes law}} \pi_t = \mathcal{B}_{Y_t}(\pi_{t|t-1})$$

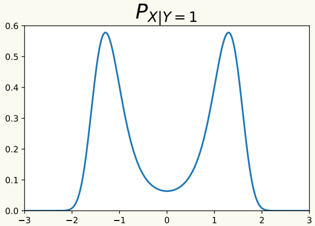
Challenge: numerical implementation of the Bayes' law

Setup:

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

EnKF:





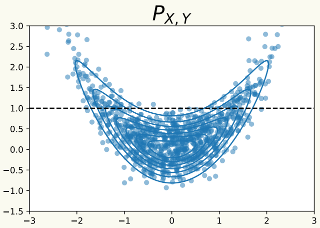
G. Evensen. "Data Assimilation. The Ensemble Kalman Filter" (2006)
S. Reich, "A dynamical systems framework for intermittent data assimilation" (2011)
E. Calvello, S. Reich, and A. M. Stuart, "Ensemble kalman methods: a mean field perspective" (2022)

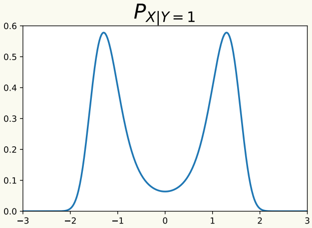
Setup:

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

EnKF:

- $(X^i, Y^i) \sim P_{X,Y}$
- fit a Gaussian
- conditioning formula for Gaussians





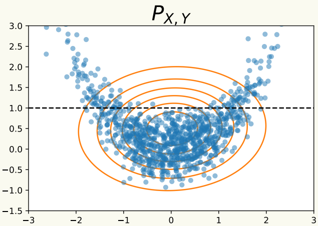
G. Evensen. "Data Assimilation. The Ensemble Kalman Filter" (2006)
S. Reich, "A dynamical systems framework for intermittent data assimilation" (2011)
E. Calvello, S. Reich, and A. M. Stuart, "Ensemble kalman methods: a mean field perspective" (2022)

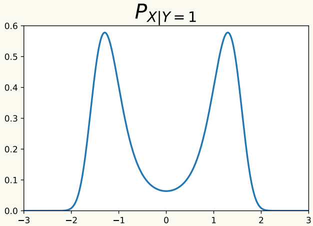
Setup:

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

EnKF:

- $(X^i, Y^i) \sim P_{X,Y}$
- fit a Gaussian
- conditioning formula for Gaussians





G. Evensen. "Data Assimilation. The Ensemble Kalman Filter" (2006)
S. Reich, "A dynamical systems framework for intermittent data assimilation" (2011)
E. Calvello, S. Reich, and A. M. Stuart, "Ensemble kalman methods: a mean field perspective" (2022)

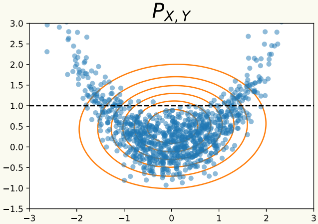
Illustrative example
Ensemble Kalman filter (EnKF)

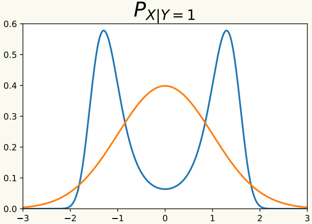
Setup:

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

EnKF:

- $(X^i, Y^i) \sim P_{X,Y}$
- fit a Gaussian
- conditioning formula for Gaussians



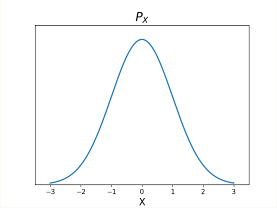


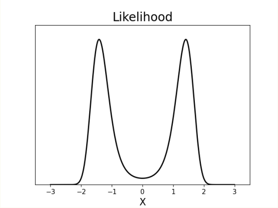
G. Evensen. "Data Assimilation. The Ensemble Kalman Filter" (2006)
S. Reich, "A dynamical systems framework for intermittent data assimilation" (2011)
E. Calvello, S. Reich, and A. M. Stuart, "Ensemble kalman methods: a mean field perspective" (2022)

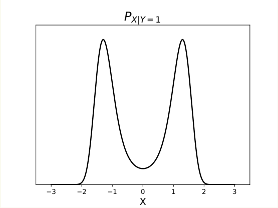
Example:

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

Importance sampling (IS):







small noise regime: $\epsilon \to 0$

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

P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, (2015)

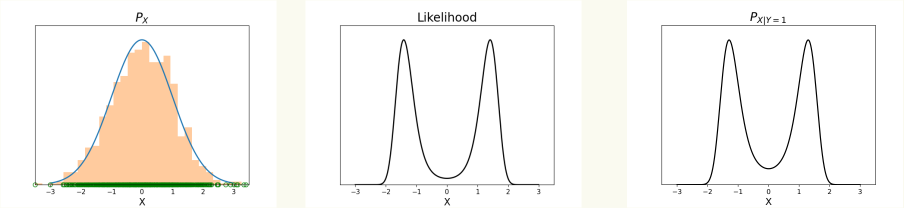
Illustrative example
Importance sampling (IS) particle filter

Example:

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

Importance sampling (IS):

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



small noise regime: $\epsilon \to 0$

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

P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, (2015)

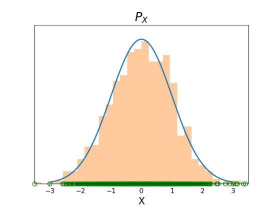
Illustrative example
Importance sampling (IS) particle filter

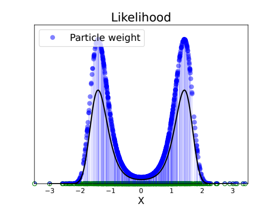
Example:

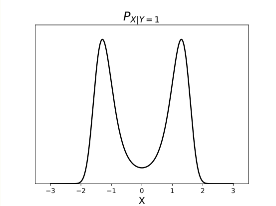
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

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

---

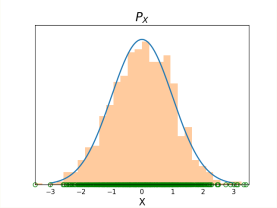
P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, (2015)

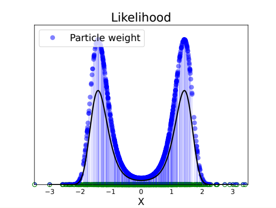
Example:

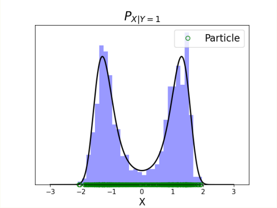
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

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

---

P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, (2015)

Illustrative example

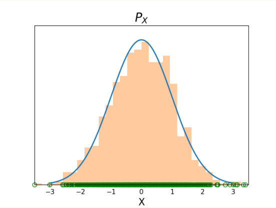
Importance sampling (IS) particle filter

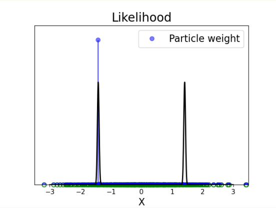
Example:

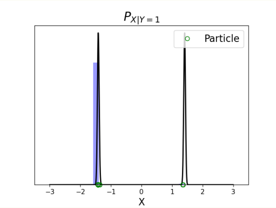
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

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

P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, (2015)

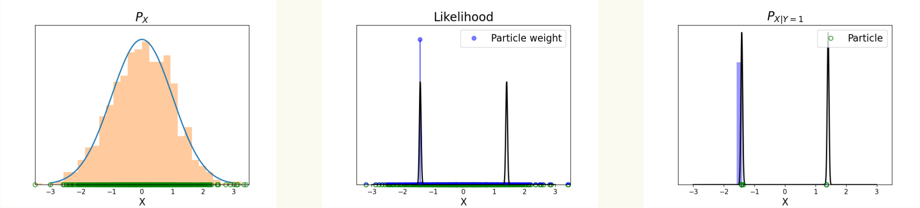
Illustrative example
Importance sampling (IS) particle filter

Example:

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

Importance sampling (IS):

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



small noise regime: $\epsilon \to 0$

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

P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, (2015)

Control and coupling techniques

- Approximate McKean-Vlasov representations [Crisan & Xiong 2010]
- Particle flow filters [Daum et. al. 2010]
- A dynamical systems framework for data assimilation [Reich. 2011]
- Feedback Particle Filter [Yang, Mehta, Meyn, 2011]
- Posterior Matching via optimal transportation [Ma & Coleman, 2011]
- Bayesian inference with optimal maps [El Moselhy & Marzouk, 2012]
- Sampling via measure transport: An introduction [Marzouk,et. al. 2016]
- Ensemble Kalman methods: a mean field perspective [Calvello et. al. 2022]
- Coupling techniques for nonlinear ensemble filtering [Spantini et. al. , 2022]
- ...

This talk: Conditioning with optimal transport map

Control and coupling techniques

- Approximate McKean-Vlasov representations [Crisan & Xiong 2010]
- Particle flow filters [Daum et. al. 2010]
- A dynamical systems framework for data assimilation [Reich. 2011]
- Feedback Particle Filter [Yang, Mehta, Meyn, 2011]
- Posterior Matching via optimal transportation [Ma & Coleman, 2011]
- Bayesian inference with optimal maps [El Moselhy & Marzouk, 2012]
- Sampling via measure transport: An introduction [Marzouk,et. al. 2016]
- Ensemble Kalman methods: a mean field perspective [Calvello et. al. 2022]
- Coupling techniques for nonlinear ensemble filtering [Spantini et. al. , 2022]
- ...

This talk: Conditioning with optimal transport map

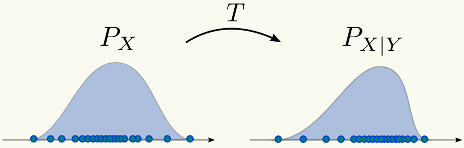
Outline

- **Part I:** Nonlinear filtering problem
- **Part II:** Optimal transport filter
- **Part III:** Extension to data-driven setting

Outline

- **Part I:** Nonlinear filtering problem
- **Part II:** Optimal transport filter
- **Part III:** Extension to data-driven setting

Conditioning with transport maps



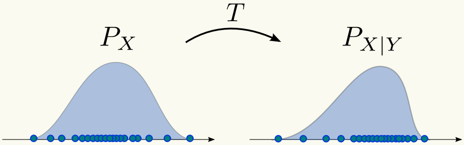
$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

Example:

Questions: In a general setting,

Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction (2016)
A. Taghvaei, P. G. Mehta. Optimal transportation methods in nonlinear filtering. (2021)
A. Spantini, R. Baptista, and Y. Marzouk. Coupling techniques for nonlinear ensemble filtering. SIAM Review (2022)

Conditioning with transport maps



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

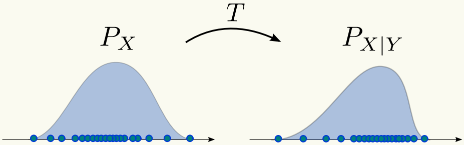
Example:

- 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: In a general setting,

Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction (2016)
A. Taghvaei, P. G. Mehta. Optimal transportation methods in nonlinear filtering. (2021)
A. Spantini, R. Baptista, and Y. Marzouk. Coupling techniques for nonlinear ensemble filtering. SIAM Review (2022)

Conditioning with transport maps



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

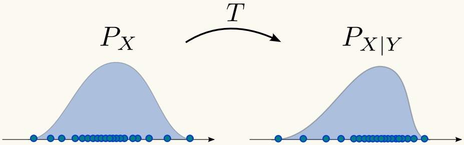
Example:

- 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: In a general setting,

Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction (2016)
A. Taghvaei, P. G. Mehta. Optimal transportation methods in nonlinear filtering. (2021)
A. Spantini, R. Baptista, and Y. Marzouk. Coupling techniques for nonlinear ensemble filtering. SIAM Review (2022)

Conditioning with transport maps



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

Example:

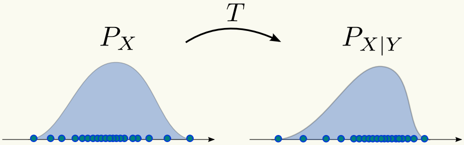
- Consider $Y = X$. Then, $P_{X|Y=y} = \delta_y$ is represented by the map $T(x, y) = y$
- Consider jointly Gaussian (X, Y). Then $P_{X|Y=y}$ is represented by the (stochastic) map $X \mapsto X + K(y - Y)$

Questions: In a general setting,

- does the map exists?
- how to numerically find it?

Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction (2016)
A. Taghvaei, P. G. Mehta. Optimal transportation methods in nonlinear filtering. (2021)
A. Spantini, R. Baptista, and Y. Marzouk. Coupling techniques for nonlinear ensemble filtering. SIAM Review (2022)

Conditioning with transport maps



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

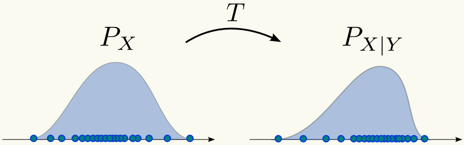
Example:

- 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: In a general setting,

- does the map exists?
- how to numerically find it?

Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction (2016)
A. Taghvaei, P. G. Mehta. Optimal transportation methods in nonlinear filtering. (2021)
A. Spantini, R. Baptista, and Y. Marzouk. Coupling techniques for nonlinear ensemble filtering. SIAM Review (2022)



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

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: In a general setting,

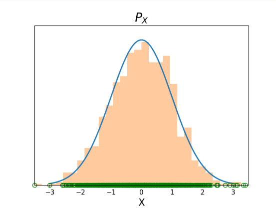
- does the map exists?
- how to numerically find it?

Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction (2016)
A. Taghvaei, P. G. Mehta. Optimal transportation methods in nonlinear filtering. (2021)
A. Spantini, R. Baptista, and Y. Marzouk. Coupling techniques for nonlinear ensemble filtering. SIAM Review (2022)

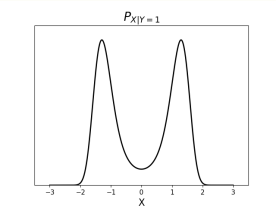
Conditioning with optimal transport map

Illustrative example

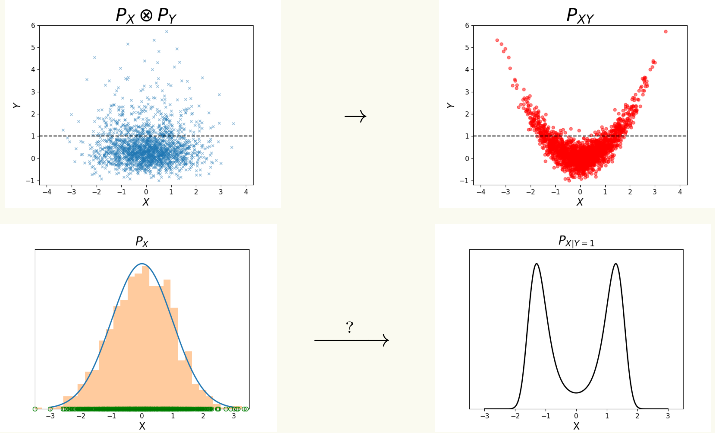
$\rightarrow$



$\xrightarrow{?}$

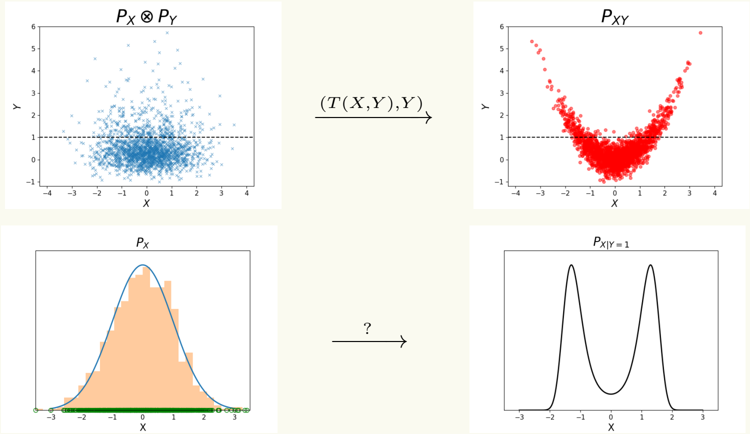


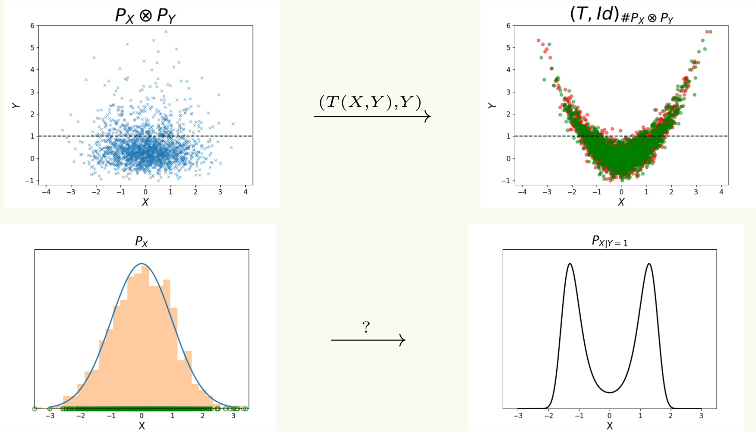
Conditioning with optimal transport map
Illustrative example



Conditioning with optimal transport map

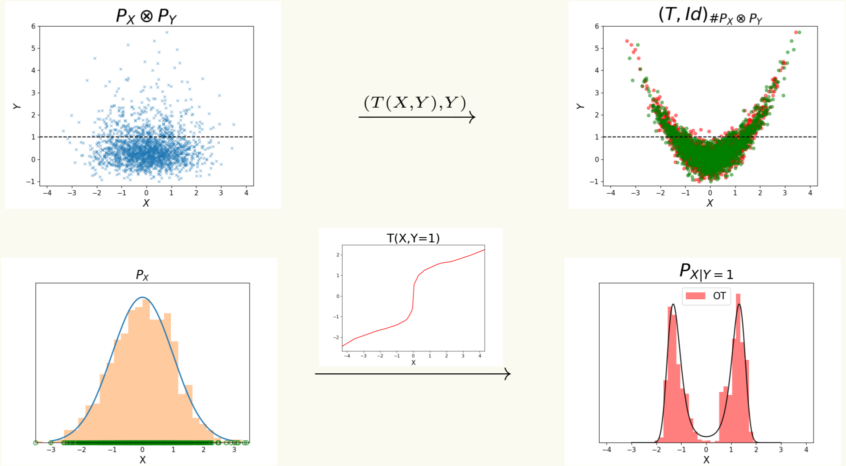
Illustrative example





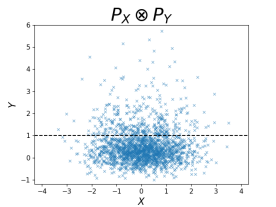
$P_X \otimes P_Y$
Y
X
$(T(X,Y),Y)$
$(T, Id)_{\#P_X \otimes P_Y}$
Y
X
P_X
X
?
$P_{X|Y=1}$
X

Conditioning with optimal transport map
Illustrative example

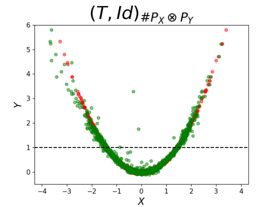


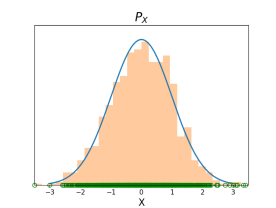
Conditioning with optimal transport map

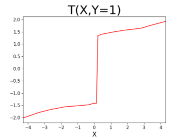
Illustrative example

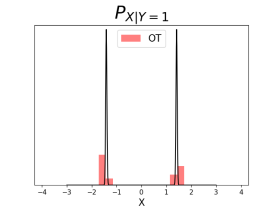


$$\xrightarrow{(T(X,Y),Y)}$$









small noise limit

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 Kantorovich semi-dual formulation:

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

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)

G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal transport approach (2016).
A. Taghvaei, and B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms (2022)
B. Hosseini, A. Hsu, A. Taghvaei. Conditional Optimal Transport on Function Spaces (2024)

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 Kantorovich semi-dual formulation:

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

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)

G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal transport approach (2016).
A. Taghvaei, and B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms (2022)
B. Hosseini, A. Hsu, A. Taghvaei. Conditional Optimal Transport on Function Spaces (2024)

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 Kantorovich semi-dual formulation:

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

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)

G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal transport approach (2016).
A. Taghvaei, and B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms (2022)
B. Hosseini, A. Hsu, A. Taghvaei. Conditional Optimal Transport on Function Spaces (2024)

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 Kantorovich semi-dual formulation:

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

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)

G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal transport approach (2016).
A. Taghvaei, and B. Hosseini, An optimal transport formulation of Bayes' law for nonlinear filtering algorithms (2022)
B. Hosseini, A. Hsu, A. Taghvaei. Conditional Optimal Transport on Function Spaces (2024)

Optimal Transport Filter

Algorithm

Initialize:

- particles $\{X_0^i\}_{i=1}^N \sim \pi_0$
- neural nets f, T

For $t = 1$ to $t = T$ do:

- propagation: $X_{t|t-1}^i \sim a(\cdot|X_{t-1}^i)$ and $Y_{t|t-1}^i \sim h(\cdot|X_{t|t-1}^i)$
- optimization: $(\hat{T}_t, \hat{f}_t) \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_{t|t-1}^i, Y_{t|t-1}^i)})$
- conditioning: $X_t^i = \hat{T}_t(X_{t|t-1}^i, Y_t)$

Remarks:

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Nonlinear Filtering with Brenier Optimal Transport Maps, ICML (2024)

Optimal Transport Filter

Algorithm

Initialize:

- particles $\{X_0^i\}_{i=1}^N \sim \pi_0$
- neural nets f, T

For $t = 1$ to $t = T$ do:

- propagation: $X_{t|t-1}^i \sim a(\cdot|X_{t-1}^i)$ and $Y_{t|t-1}^i \sim h(\cdot|X_{t|t-1}^i)$
- optimization: $(\hat{T}_t, \hat{f}_t) \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N} \sum_{i=1}^N \delta_{(X_{t|t-1}^i, Y_{t|t-1}^i)})$
- conditioning: $X_t^i = \hat{T}_t(X_{t|t-1}^i, Y_t)$

Remarks:

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Nonlinear Filtering with Brenier Optimal Transport Maps, ICML (2024)

Optimal Transport Filter

Algorithm

Initialize:

- particles $\{X_0^i\}_{i=1}^N \sim \pi_0$
- neural nets f, T

For $t = 1$ to $t = T$ do:

- propagation: $X_{t|t-1}^i \sim a(\cdot|X_{t-1}^i)$ and $Y_{t|t-1}^i \sim h(\cdot|X_{t|t-1}^i)$
- optimization: $(\hat{T}_t, \hat{f}_t) \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N} \sum_{i=1}^N \delta_{(X_{t|t-1}^i, Y_{t|t-1}^i)})$
- conditioning: $X_t^i = \hat{T}_t(X_{t|t-1}^i, Y_t)$

Remarks:

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Nonlinear Filtering with Brenier Optimal Transport Maps, ICML (2024)

Optimal Transport Filter

Algorithm

Initialize:

- particles $\{X_0^i\}_{i=1}^N \sim \pi_0$
- neural nets f, T

For $t = 1$ to $t = T$ do:

- propagation: $X_{t|t-1}^i \sim a(\cdot|X_{t-1}^i)$ and $Y_{t|t-1}^i \sim h(\cdot|X_{t|t-1}^i)$
- optimization: $(\hat{T}_t, \hat{f}_t) \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_{t|t-1}^i, Y_{t|t-1}^i)})$
- conditioning: $X_t^i = \hat{T}_t(X_{t|t-1}^i, Y_t)$

Remarks:

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Nonlinear Filtering with Brenier Optimal Transport Maps, ICML (2024)

Optimal Transport Filter

Algorithm

Initialize:

- particles $\{X_0^i\}_{i=1}^N \sim \pi_0$
- neural nets f, T

For $t = 1$ to $t = T$ do:

- propagation: $X_{t|t-1}^i \sim a(\cdot|X_{t-1}^i)$ and $Y_{t|t-1}^i \sim h(\cdot|X_{t|t-1}^i)$
- optimization: $(\hat{T}_t, \hat{f}_t) \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N} \sum_{i=1}^N \delta_{(X_{t|t-1}^i, Y_{t|t-1}^i)})$
- conditioning: $X_t^i = \hat{T}_t(X_{t|t-1}^i, Y_t)$

Remarks:

- No need for analytical form of the dynamic and observation models
- In practice, only a few iterations of the optimization is performed

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Nonlinear Filtering with Brenier Optimal Transport Maps, ICML (2024)

Algorithm

Initialize:

- particles $\{X_0^i\}_{i=1}^N \sim \pi_0$
- neural nets f, T

For $t = 1$ to $t = T$ do:

- propagation: $X_{t|t-1}^i \sim a(\cdot|X_{t-1}^i)$ and $Y_{t|t-1}^i \sim h(\cdot|X_{t|t-1}^i)$
- optimization: $(\hat{T}_t, \hat{f}_t) \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X_{t|t-1}^i, Y_{t|t-1}^i)})$
- conditioning: $X_t^i = \hat{T}_t(X_{t|t-1}^i, Y_t)$

Remarks:

- No need for analytical form of the dynamic and observation models
- In practice, only a few iterations of the optimization is performed

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, Nonlinear Filtering with Brenier Optimal Transport Maps, ICML (2024)

Outline

- **Part I:** Nonlinear filtering problem
- **Part II:** Optimal transport filter
- **Part III:** Extension to data-driven setting

Outline

- **Part I:** Nonlinear filtering problem
- **Part II:** Optimal transport filter
- **Part III:** Extension to data-driven setting

Problem setup:

$$X_t \sim a(\cdot \mid X_{t-1}), \quad X_0 \sim \pi_0$$
$$Y_t \sim h(\cdot \mid X_t)$$

- X_t is the state
- Y_t is the observation
- the dynamic and observation models are unknown

Objective:

given: $\{X_0^j, (X_1^j, Y_1^j), \ldots, (X_{t_f}^j, Y_{t_f}^j)\}_{j=1}^J$

compute: $\pi_t := P(X_t | Y_t, \ldots, Y_1), \quad \forall t \geq 0$

for a new set of observations $\{Y_t, \ldots, Y_1\}$

Nonlinear filtering problem
Data-driven setting

Problem setup:

$$X_t \sim a(\cdot \mid X_{t-1}), \quad X_0 \sim \pi_0$$
$$Y_t \sim h(\cdot \mid X_t)$$

- X_t is the state
- Y_t is the observation
- the dynamic and observation models are unknown

Objective:

given: $\{X_0^j, (X_1^j, Y_1^j), \ldots, (X_{t_f}^j, Y_{t_f}^j)\}_{j=1}^J$

compute: $\pi_t := P(X_t | Y_t, \ldots, Y_1), \quad \forall t \geq 0$

for a new set of observations $\{Y_t, \ldots, Y_1\}$

Solution approach

- Exact posterior:

$$\pi_t := \mathbb{P}_{X_0 \sim \pi_0}(X_t | Y_t, \ldots, Y_1)$$

- Step 1: Truncated posterior

$$\pi^{\mu}_{t,s} := \mathbb{P}_{X_s \sim \mu}(X_t | Y_t, \ldots, Y_{s+1})$$

- Step 2: OT representation

$$\pi^{\mu}_{t,s} = T(\cdot, Y_t, \ldots, Y_s) \# \mu \quad \text{where}$$
$$T \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; P_{X_t, Y_t, \ldots, Y_{s+1}})$$

- Step 3: Stationary assumption

$$P_{X_t, Y_t, \ldots, Y_{s+1}} = P_{X_w, Y_w, \ldots, Y_1} \quad \text{where} \quad w := t - s$$

- Step 4: Use training data to approximate $P_{X_w, Y_w, \ldots, Y_1}$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Solution approach

- Exact posterior:

$$\pi_t := \mathbb{P}_{X_0 \sim \pi_0}(X_t | Y_t, \ldots, Y_1)$$

- Step 1: Truncated posterior

$$\pi^{\mu}_{t,s} := \mathbb{P}_{X_s \sim \mu}(X_t | Y_t, \ldots, Y_{s+1})$$

- Step 2: OT representation

$$\pi^{\mu}_{t,s} = T(\cdot, Y_t, \ldots, Y_s) \# \mu \quad \text{where}$$
$$T \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; P_{X_t, Y_t, \ldots, Y_{s+1}})$$

- Step 3: Stationary assumption

$$P_{X_t, Y_t, \ldots, Y_{s+1}} = P_{X_w, Y_w, \ldots, Y_1} \quad \text{where} \quad w := t - s$$

- Step 4: Use training data to approximate $P_{X_w, Y_w, \ldots, Y_1}$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Solution approach

- Exact posterior:

$$\pi_t := \mathbb{P}_{X_0 \sim \pi_0}(X_t | Y_t, \ldots, Y_1)$$

- Step 1: Truncated posterior

$$\pi_{t,s}^{\mu} := \mathbb{P}_{X_s \sim \mu}(X_t | Y_t, \ldots, Y_{s+1})$$

- Step 2: OT representation

$$\pi_{t,s}^{\mu} = T(\cdot, Y_t, \ldots, Y_s) \# \mu \quad \text{where}$$
$$T \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; P_{X_t, Y_t, \ldots, Y_{s+1}})$$

- Step 3: Stationary assumption

$$P_{X_t, Y_t, \ldots, Y_{s+1}} = P_{X_w, Y_w, \ldots, Y_1} \quad \text{where} \quad w := t - s$$

- Step 4: Use training data to approximate $P_{X_w, Y_w, \ldots, Y_1}$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Solution approach

- Exact posterior:

$$\pi_t := \mathbb{P}_{X_0 \sim \pi_0}(X_t | Y_t, \ldots, Y_1)$$

- Step 1: Truncated posterior

$$\pi_{t,s}^{\mu} := \mathbb{P}_{X_s \sim \mu}(X_t | Y_t, \ldots, Y_{s+1})$$

- Step 2: OT representation

$$\pi_{t,s}^{\mu} = T(\cdot, Y_t, \ldots, Y_s) \# \mu \quad \text{where}$$
$$T \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; P_{X_t, Y_t, \ldots, Y_{s+1}})$$

- Step 3: Stationary assumption

$$P_{X_t, Y_t, \ldots, Y_{s+1}} = P_{X_w, Y_w, \ldots, Y_1} \quad \text{where} \quad w := t - s$$

- Step 4: Use training data to approximate $P_{X_w, Y_w, \ldots, Y_1}$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Solution approach

- Exact posterior:

$$\pi_t := \mathbb{P}_{X_0 \sim \pi_0}(X_t | Y_t, \ldots, Y_1)$$

- Step 1: Truncated posterior

$$\pi^{\mu}_{t,s} := \mathbb{P}_{X_s \sim \mu}(X_t | Y_t, \ldots, Y_{s+1})$$

- Step 2: OT representation

$$\pi^{\mu}_{t,s} = T(\cdot, Y_t, \ldots, Y_s)\#\mu \quad \text{where}$$
$$T \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; P_{X_t, Y_t, \ldots, Y_{s+1}})$$

- Step 3: Stationary assumption

$$P_{X_t, Y_t, \ldots, Y_{s+1}} = P_{X_w, Y_w, \ldots, Y_1} \quad \text{where} \quad w := t - s$$

- Step 4: Use training data to approximate $P_{X_w, Y_w, \ldots, Y_1}$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Optimal transport data-driven filter (OT-DDF)

Offline stage:

- **Input:** Recorded data $\{X_0^j, (X_1^j, Y_1^j), \ldots, (X_w^j, Y_w^j)\}_{j=1}^J$
- Obtain map $\hat{T}_w$ by solving

$$\max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T, \text{Data})$$

- **Output:** The map $\hat{T}_w$

Online stage:

Optimal transport data-driven filter (OT-DDF)

Offline stage:

- **Input:** Recorded data $\{X_0^j, (X_1^j, Y_1^j), \ldots, (X_w^j, Y_w^j)\}_{j=1}^J$
- Obtain map $\hat{T}_w$ by solving

$$\max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T, \text{Data})$$

- **Output:** The map $\hat{T}_w$

Online stage:

- **Input:** Map $\hat{T}_w$ and initial particles $X_0^i \sim \pi_0$
- Update particles $X_t^i = \hat{T}_w(X_0^i, Y_t, Y_{t-1}, \ldots, Y_{t-w})$
- **Output:** Particles $\{X_t^i\}_{i=1}^N, \quad \forall\, t \geq w+1$

Numerical example: Dynamical model

$$X_t = \begin{bmatrix} \alpha & \sqrt{1-\alpha^2} \\ -\sqrt{1-\alpha^2} & \alpha \end{bmatrix} X_{t-1} + \sigma V_t, \quad \alpha = 0.9, \sigma = 0.1$$

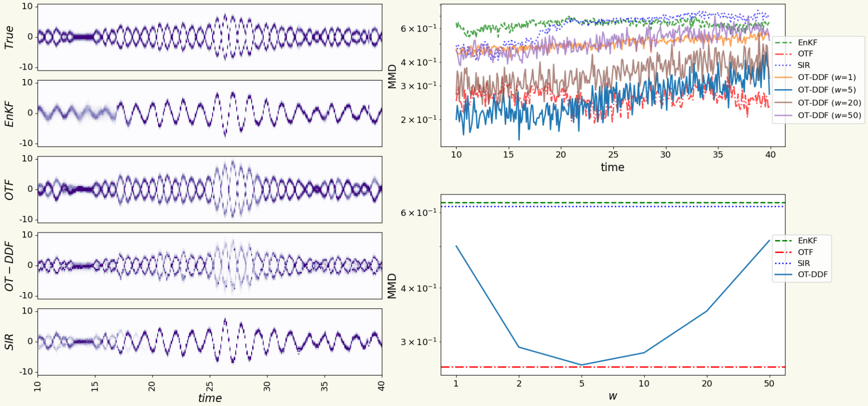
$$Y_t = h(X_t) + \sigma W_t$$

Numerical example: Dynamical model

$$X_t = \begin{bmatrix} \alpha & \sqrt{1-\alpha^2} \\ -\sqrt{1-\alpha^2} & \alpha \end{bmatrix} X_{t-1} + \sigma V_t, \quad \alpha = 0.9, \sigma = 0.1$$

$$Y_t = h(X_t) + \sigma W_t$$

$$h(X_t) = X_t^2(1)$$



Numerical example: Lorenz 63

$$\begin{bmatrix} \dot{X}(1) \\ \dot{X}(2) \\ \dot{X}(3) \end{bmatrix} = \begin{bmatrix} \sigma(X(2) - X(1)) \\ X(1)(\rho - X(3)) - X(2) \\ X(1)X(2) - \beta X(3) \end{bmatrix}, \quad X_0 \sim \mathcal{N}(0, 10 \cdot I_3),$$

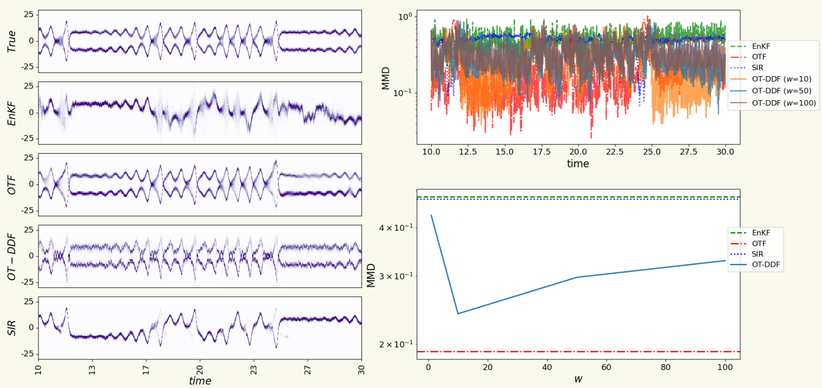
$$Y_t = h(X_t) + \sigma W_t, \quad \sigma^2 = 1$$

Numerical example: Lorenz 63

$$\begin{bmatrix} \dot{X}(1) \\ \dot{X}(2) \\ \dot{X}(3) \end{bmatrix} = \begin{bmatrix} \sigma(X(2) - X(1)) \\ X(1)(\rho - X(3)) - X(2) \\ X(1)X(2) - \beta X(3) \end{bmatrix}, \quad X_0 \sim \mathcal{N}(0, 10 \cdot I_3),$$

$$Y_t = h(X_t) + \sigma W_t, \quad \sigma^2 = 1$$

$$h(X_t) = X_t(3)$$



Numerical example: Lorenz 63

$$\begin{bmatrix}\dot X(1)\\ \dot X(2)\\ \dot X(3)\end{bmatrix} = \begin{bmatrix}\sigma(X(2)-X(1))\\ X(1)(\rho - X(3)) - X(2)\\ X(1)X(2) - \beta X(3)\end{bmatrix}, \quad X_0 \sim \mathcal{N}(0, 10\cdot I_3),$$

$$Y_t = h(X_t) + \sigma W_t, \quad \sigma^2 = 1$$

Offline training time: 46.29 seconds

One-time step update:

Method	EnKF	SIR	OTF	OT-DDF
time(sec)	1.7×10^{-4}	2.0×10^{-4}	6.8×10^{-2}	1.5×10^{-4}

Error analysis

Main result

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \ldots, Y_{t-w})\#\mu, \pi_t) \leq C\lambda^w M + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Error analysis

Main result

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \ldots, Y_{t-w})\#\mu, \pi_t) \leq C\lambda^w M + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Error analysis

Main result

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \ldots, Y_{t-w})\#\mu, \pi_t) \leq C\lambda^w M + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Error analysis

Main result

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \ldots, Y_{t-w})\#\mu, \pi_t) \leq C\lambda^w M + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Error analysis

Main result

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \ldots, Y_{t-w})\#\mu, \pi_t) \le C\lambda^w M + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Error analysis

Main result

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \ldots, Y_{t-w})\#\mu, \pi_t) \leq C\lambda^w M + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps (2024)

Acknowledgments

M. Al-Jarrah

N. Jin

B. Hosseini

NSF

References: