Generative Models for Nonlinear Filtering: A Conditional Optimal Transport Approach

Presented at the
Univ. of Washington ECE Research Colloquium Series

Amirhossein Taghvaei

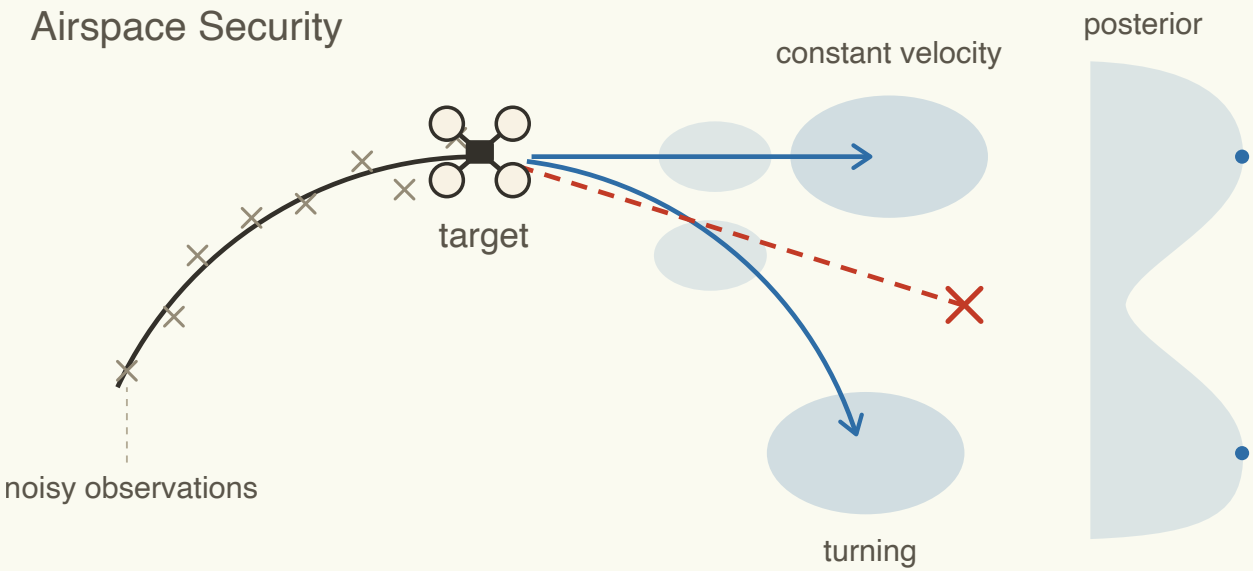
Department of Aeronautics & Astronautics
University of Washington, Seattle

Oct 7, 2026



Research overview

Interplay of filtering and generative models



Classical filters

- physics-based
- uni-modal uncertainty
- not scalable to high dim.

Our research

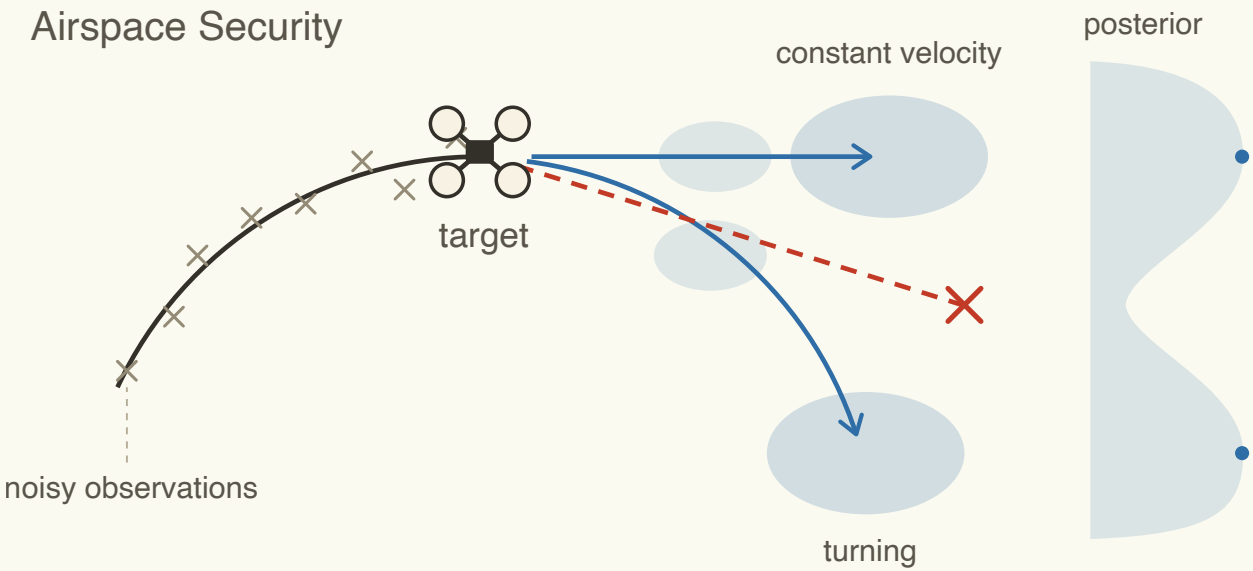
Conditional OT

Generative models

- data-driven
- multi-modal uncertainty
- scalable to high dim.

Research overview

Interplay of filtering and generative models



Classical filters

- physics-based
- uni-modal uncertainty
- not scalable to high dim.

Our research

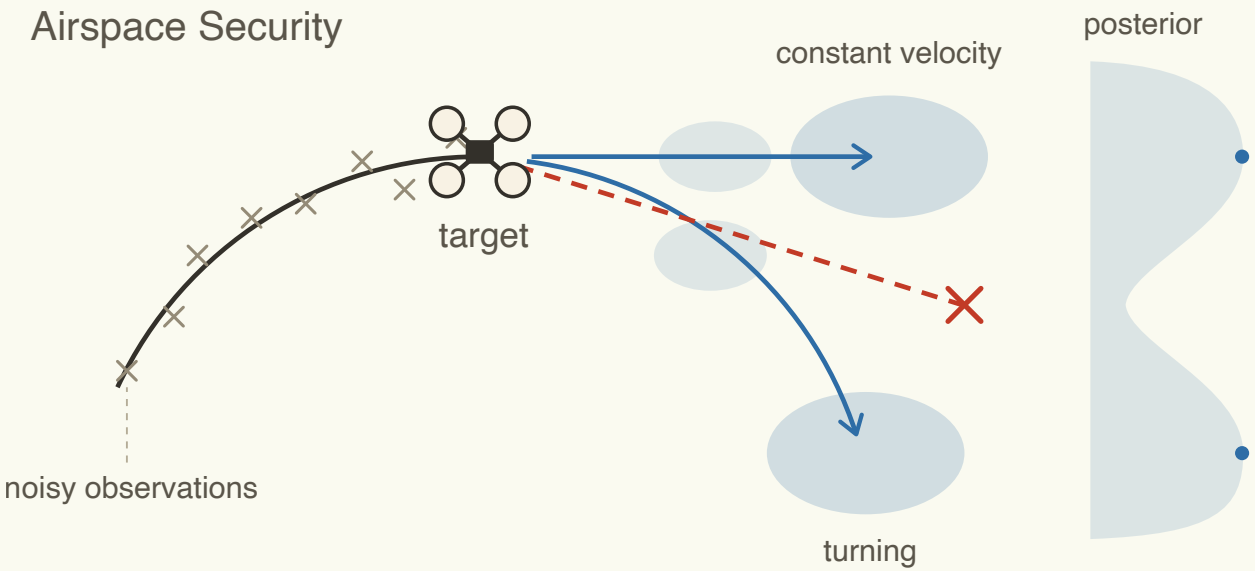
Conditional OT

Generative models

- data-driven
- multi-modal uncertainty
- scalable to high dim.

Research overview

Interplay of filtering and generative models



Classical filters

- physics-based
- uni-modal uncertainty
- not scalable to high dim.

Our research

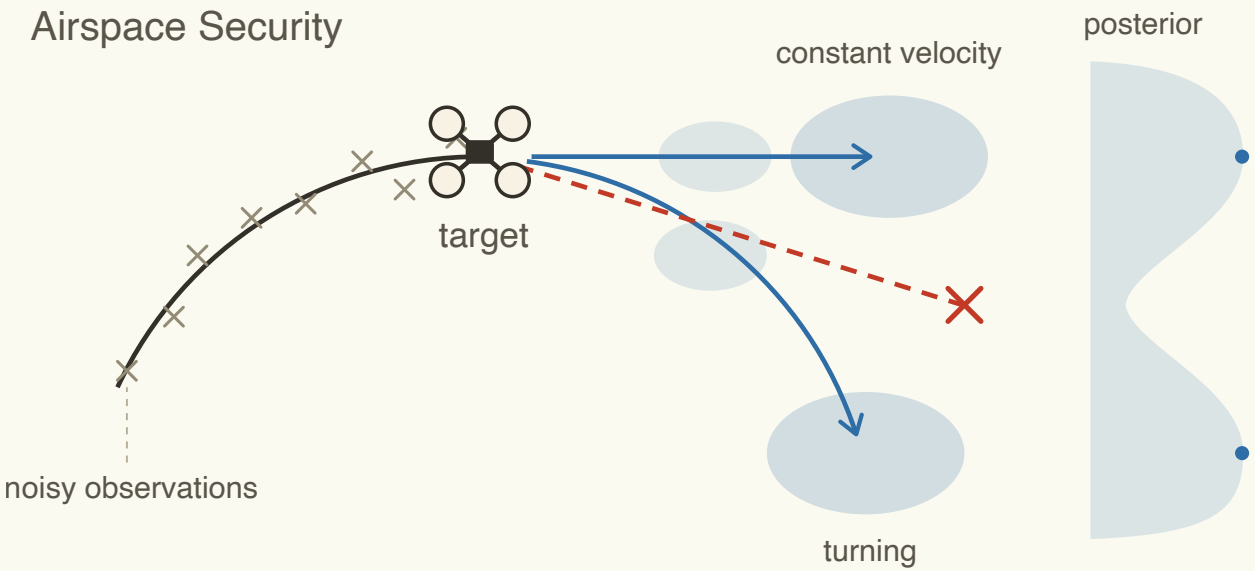
Conditional OT

Generative models

- data-driven
- multi-modal uncertainty
- scalable to high dim.

Research overview

Interplay of filtering and generative models



Classical filters

- physics-based
- uni-modal uncertainty
- not scalable to high dim.

Our research

Conditional OT

Generative models

- data-driven
- multi-modal uncertainty
- scalable to high dim.

Research overview
Highlights

Foundation:

- Optimal Transport Formulation of the EnKF (TAC 2021)
- Optimal Transport Methods in Nonlinear Filtering (CSM 2021)
- Optimal transport formulation of the Bayes' law (CDC 2022)

Algorithm development:

- Extension to manifold for attitude estimation (L-CSS 2023)
- full algorithm + benchmarks, beats SOTA (ICML 2024)
- Data-driven variant (CDC 2024, Al-Jarrah wins outstanding student paper award)
- Fast filtering with amortization (L-CSS 2025)

Theoretical development:

- Conditional Optimal Transport (SIAM UQ 2025)
- Statistical error analysis (2nd round of review SIAM UQ)

Funding:

- National Science Foundation, Energy Power Control Networks (EPCN)
- AFOSR Young Investigator Program (YIP)

Research overview
Highlights

Foundation:

- Optimal Transport Formulation of the EnKF (TAC 2021)
- Optimal Transport Methods in Nonlinear Filtering (CSM 2021)
- Optimal transport formulation of the Bayes' law (CDC 2022)

Algorithm development:

- Extension to manifold for attitude estimation (L-CSS 2023)
- full algorithm + benchmarks, beats SOTA (ICML 2024)
- Data-driven variant (CDC 2024, Al-Jarrah wins outstanding student paper award)
- Fast filtering with amortization (L-CSS 2025)

Theoretical development:

- Conditional Optimal Transport (SIAM UQ 2025)
- Statistical error analysis (2nd round of review SIAM UQ)

Funding:

- National Science Foundation, Energy Power Control Networks (EPCN)
- AFOSR Young Investigator Program (YIP)

Research overview
Highlights

Foundation:

- Optimal Transport Formulation of the EnKF (TAC 2021)
- Optimal Transport Methods in Nonlinear Filtering (CSM 2021)
- Optimal transport formulation of the Bayes' law (CDC 2022)

Algorithm development:

- Extension to manifold for attitude estimation (L-CSS 2023)
- full algorithm + benchmarks, beats SOTA (ICML 2024)
- Data-driven variant (CDC 2024, Al-Jarrah wins outstanding student paper award)
- Fast filtering with amortization (L-CSS 2025)

Theoretical development:

- Conditional Optimal Transport (SIAM UQ 2025)
- Statistical error analysis (2nd round of review SIAM UQ)

Funding:

- National Science Foundation, Energy Power Control Networks (EPCN)
- AFOSR Young Investigator Program (YIP)

Research overview
Highlights

Foundation:

- Optimal Transport Formulation of the EnKF (TAC 2021)
- Optimal Transport Methods in Nonlinear Filtering (CSM 2021)
- Optimal transport formulation of the Bayes' law (CDC 2022)

Algorithm development:

- Extension to manifold for attitude estimation (L-CSS 2023)
- full algorithm + benchmarks, beats SOTA (ICML 2024)
- Data-driven variant (CDC 2024, Al-Jarrah wins outstanding student paper award)
- Fast filtering with amortization (L-CSS 2025)

Theoretical development:

- Conditional Optimal Transport (SIAM UQ 2025)
- Statistical error analysis (2nd round of review SIAM UQ)

Funding:

- National Science Foundation, Energy Power Control Networks (EPCN)
- AFOSR Young Investigator Program (YIP)

Outline

- Foundations
  - Bayes' law and its fundamental challenges
  - Conditioning with optimal transport maps
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Outline

- Foundations
  - Bayes' law and its fundamental challenges
  - Conditioning with optimal transport maps
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Data-driven implementation of Bayes' law

Problem:

- Hidden state X, noisy observation Y
- What is the conditional probability distribution of X given Y? (posterior)

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

Data-driven setting:

- $P_{X,Y}$ is not available.

$$\textbf{Given:} \quad (X^i, Y^i)_{i=1}^N \overset{\text{i.i.d.}}{\sim} P_{X,Y}$$
$$\textbf{Goal:} \quad \text{generate samples from } P_{X|Y=y} \text{ for } \textit{any } y$$

Data-driven implementation of Bayes' law

Problem:

- Hidden state X, noisy observation Y
- What is the conditional probability distribution of X given Y? (posterior)

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

Data-driven setting:

- $P_{X,Y}$ is not available.

Given: $(X^i, Y^i)_{i=1}^N \overset{\text{i.i.d.}}{\sim} P_{X,Y}$

Goal: generate samples from $P_{X|Y=y}$ for *any* y

Data-driven implementation of Bayes' law

Problem:

- Hidden state X, noisy observation Y
- What is the conditional probability distribution of X given Y? (posterior)

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

Data-driven setting:

- $P_{X,Y}$ is not available.

Given: $(X^i, Y^i)_{i=1}^N \overset{\text{i.i.d.}}{\sim} P_{X,Y}$

Goal: generate samples from $P_{X|Y=y}$ for *any* y

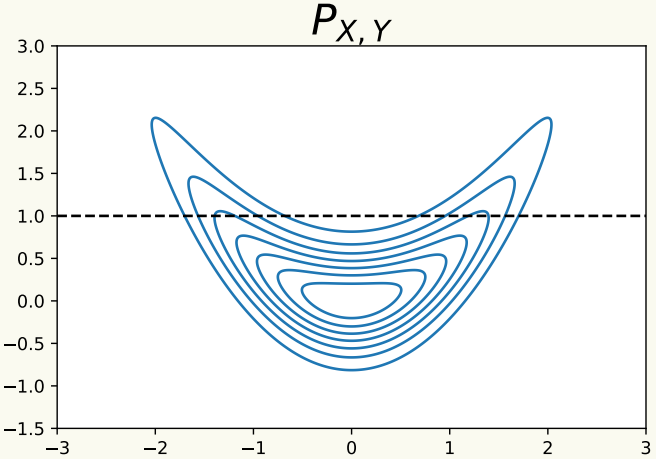
Limitations of the existing solutions
Ensemble Kalman filter (EnKF)

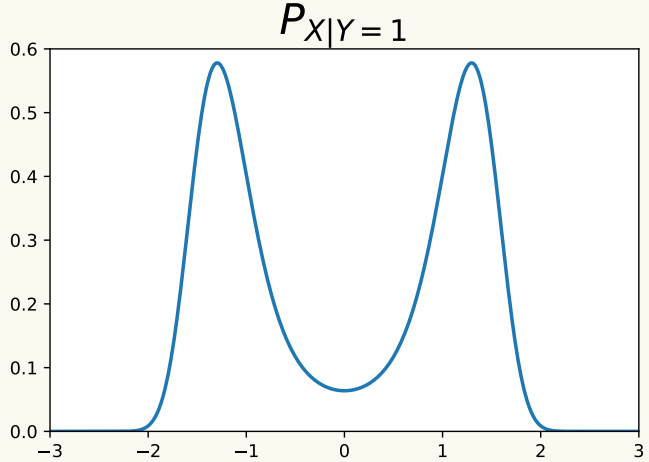
Setup:

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

EnKF:

- $[X, Y] \sim P_{X,Y}$
- 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)

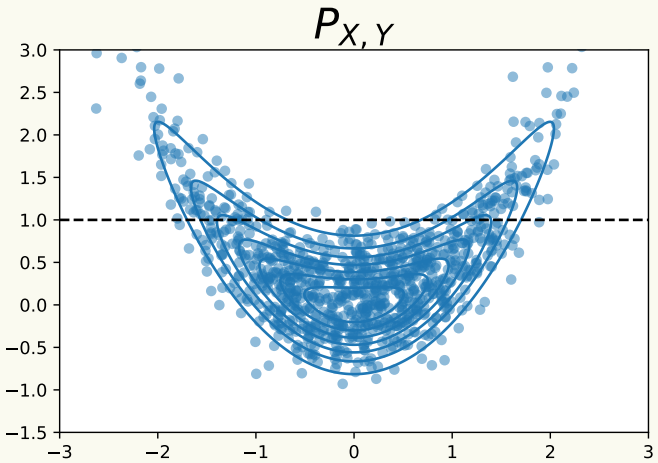
Limitations of the existing solutions
Ensemble Kalman filter (EnKF)

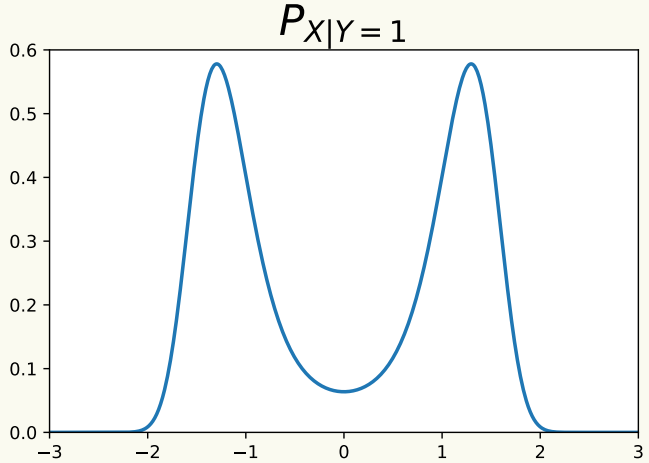
Setup:

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

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)

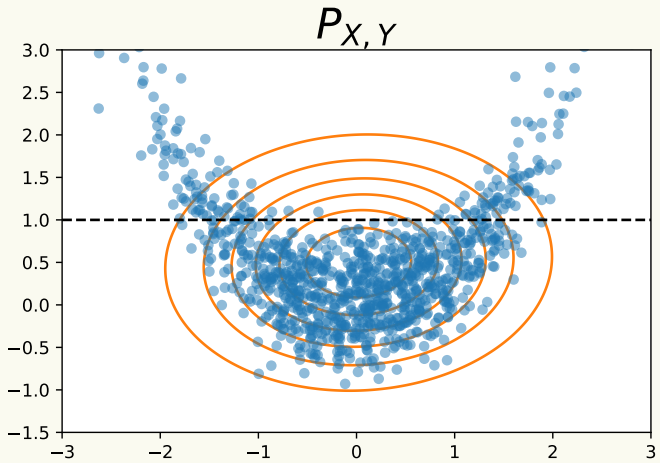
Limitations of the existing solutions
Ensemble Kalman filter (EnKF)

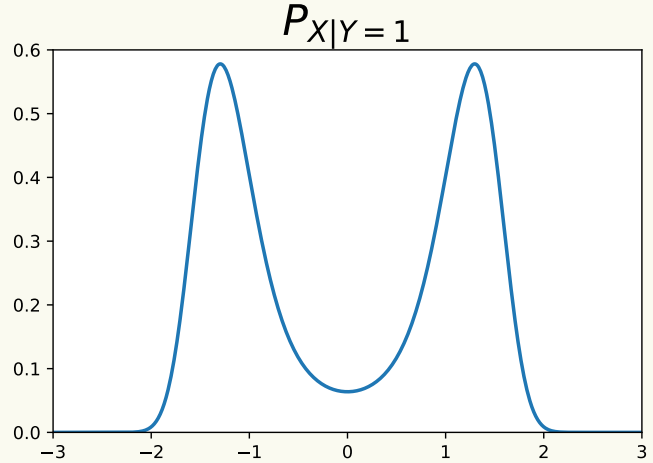
Setup:

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

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)

Limitations of the existing solutions

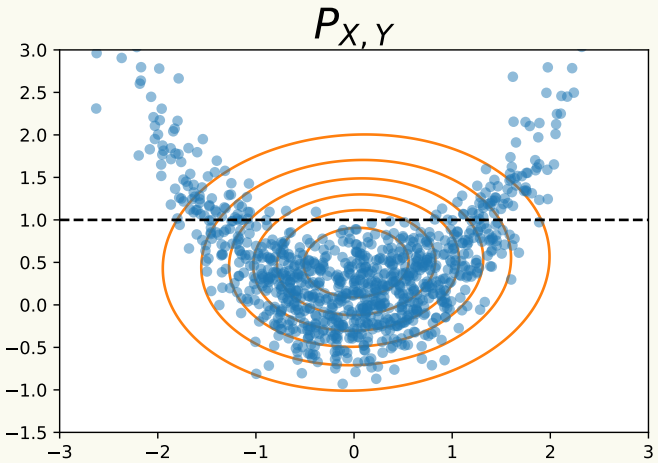
Ensemble Kalman filter (EnKF)

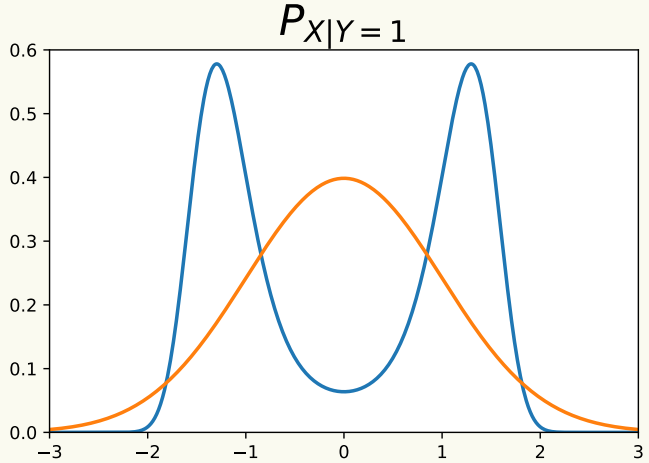
Setup:

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

EnKF:

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





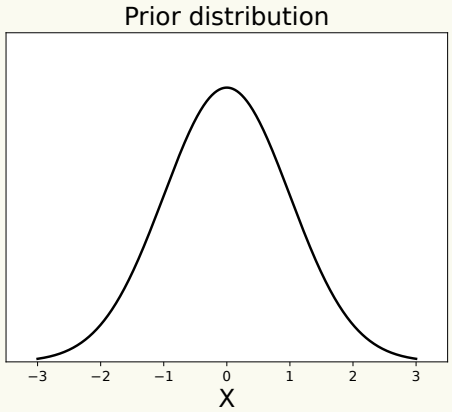
G. Evensen. "Data Assimilation. The Ensemble Kalman Filter" (2006)
S. Reich, "A dynamical systems framework for intermittent data assimilation" (2011)
E. Calvello, S. Reich, and A. M. Stuart, "Ensemble kalman methods: a mean field perspective" (2022)

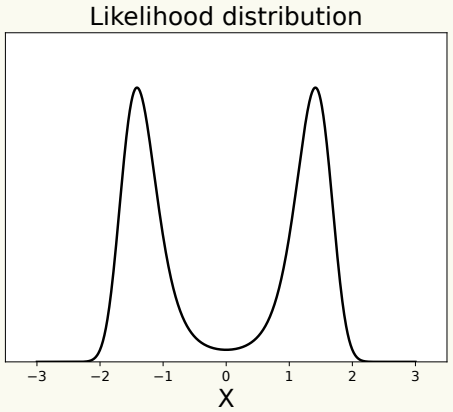
Limitations of the existing solutions

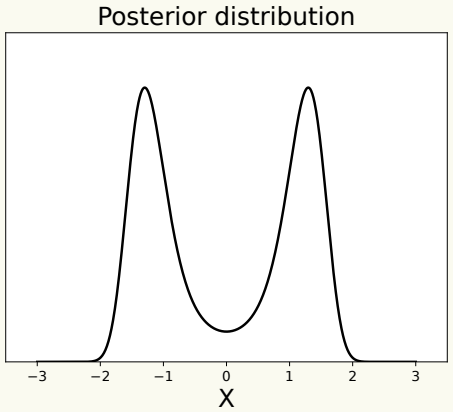
Importance sampling (IS) particle filter

Example:

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







Importance sampling (IS):

- $X^{(i)} \sim \mathcal{N}(0,1)$
- $w^{(i)} \propto \frac{1}{\sqrt{2\pi\epsilon}} \exp\left(-\frac{(1-\frac{1}{2}(X^{(i)})^2)^2}{2\epsilon}\right)$
- $P_{X|Y=1} \approx \sum_{i=1}^{N} w^{(i)} \delta_{X^{(i)}}$

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)

Limitations of the existing solutions

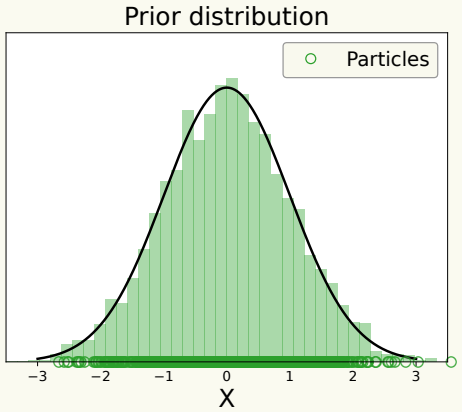
Importance sampling (IS) particle filter

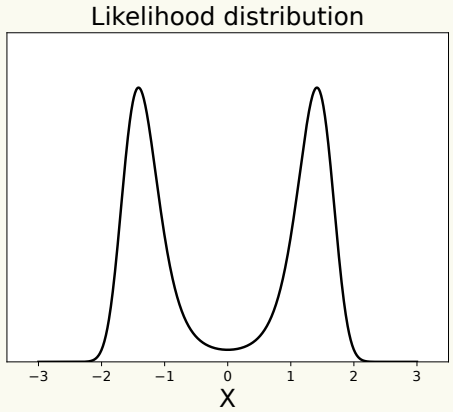
Example:

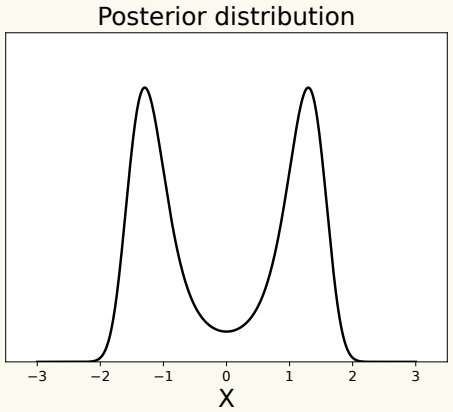
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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)

Limitations of the existing solutions

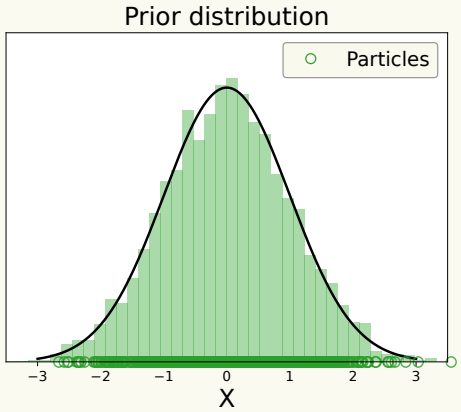
Importance sampling (IS) particle filter

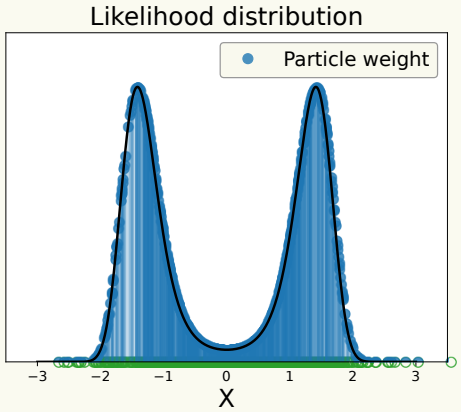
Example:

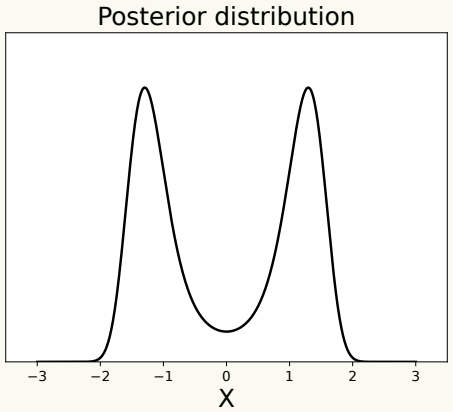
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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)

Limitations of the existing solutions

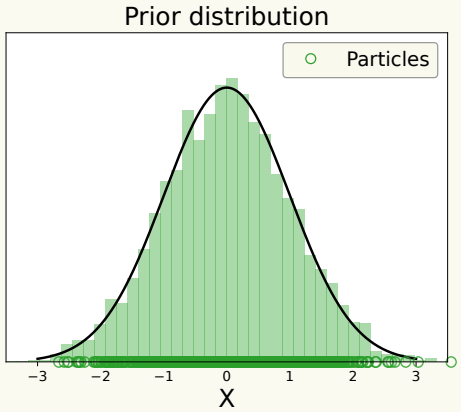
Importance sampling (IS) particle filter

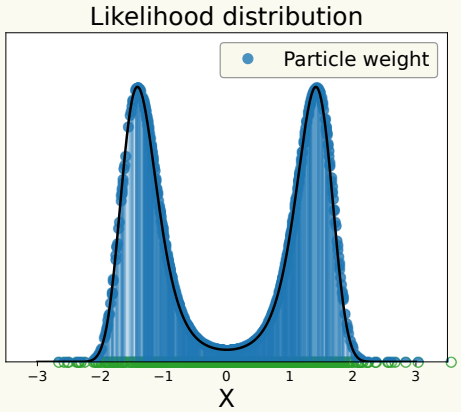
Example:

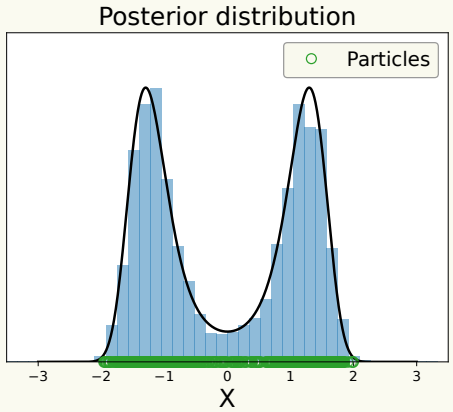
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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)

Limitations of the existing solutions

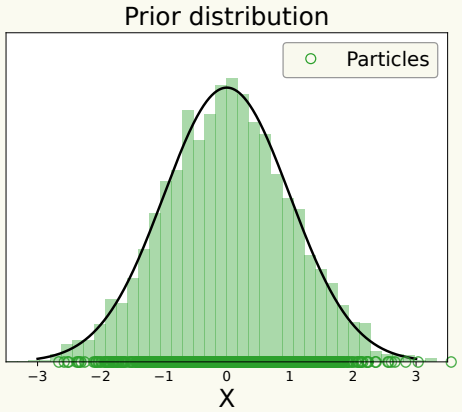
Importance sampling (IS) particle filter

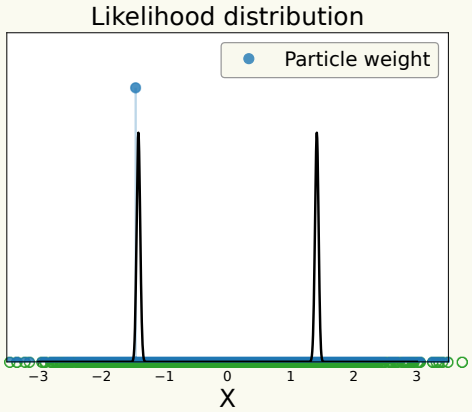
Example:

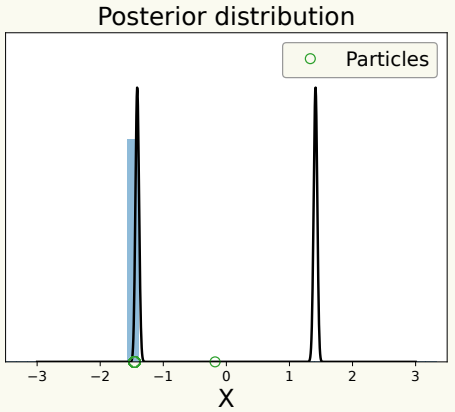
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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)

Limitations of the existing solutions

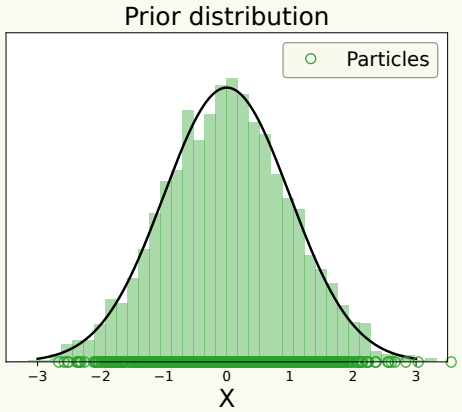
Importance sampling (IS) particle filter

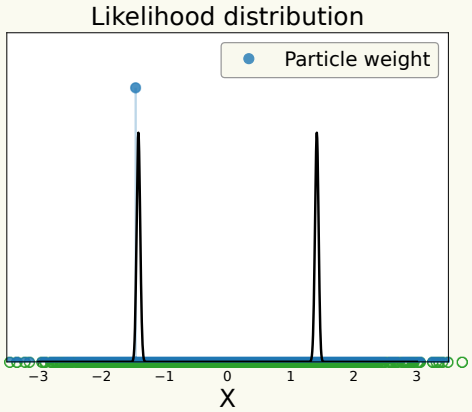
Example:

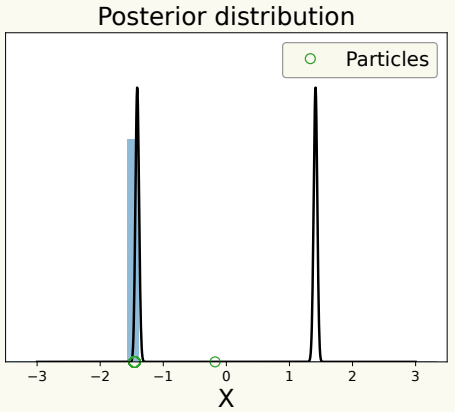
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

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 or dynamical system techniques

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

Our approach: Conditioning with optimal transport map

Control or dynamical system techniques

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

Our approach: Conditioning with optimal transport map

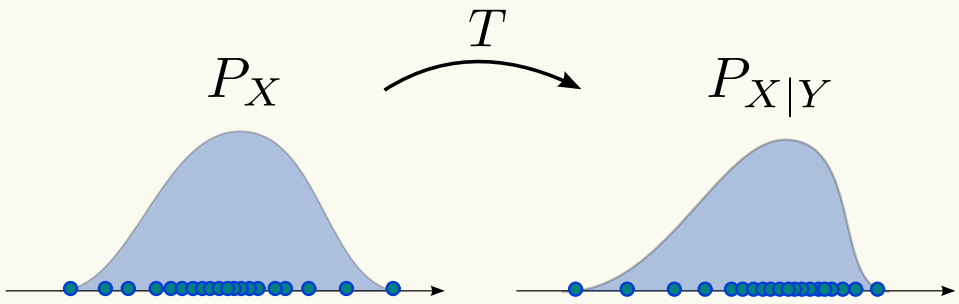
Outline

- Foundations
  - Bayes' law and its fundamental challenges
  - Conditioning with optimal transport maps
- Algorithmic development

Outline

- Foundations
  - Bayes' law and its fundamental challenges
  - Conditioning with optimal transport maps
- Algorithmic development

Conditioning with transport maps



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

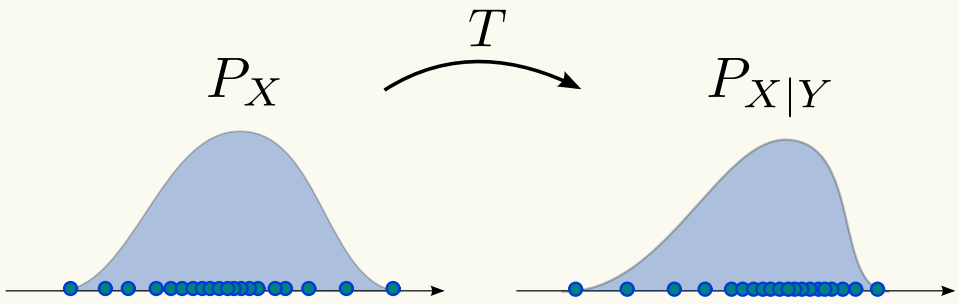
Example:

- Consider $Y = X$. Then $P_{X|Y=y}$ is represented by the map $T(x, y) = y$.
- Consider jointly Gaussian (X, Y). Then $P_{X|Y=y}$ is represented by the (autoencoder) map $T(x, y) = A(y - b) + x$.

Questions: In a general setting,

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

Conditioning with transport maps



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

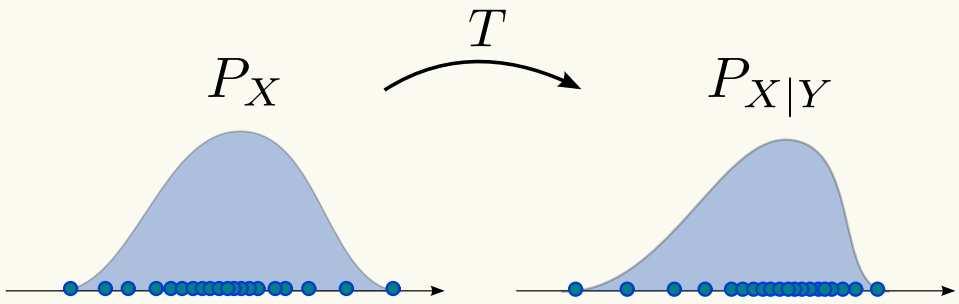
Example:

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

Questions: In a general setting,

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

Conditioning with transport maps



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

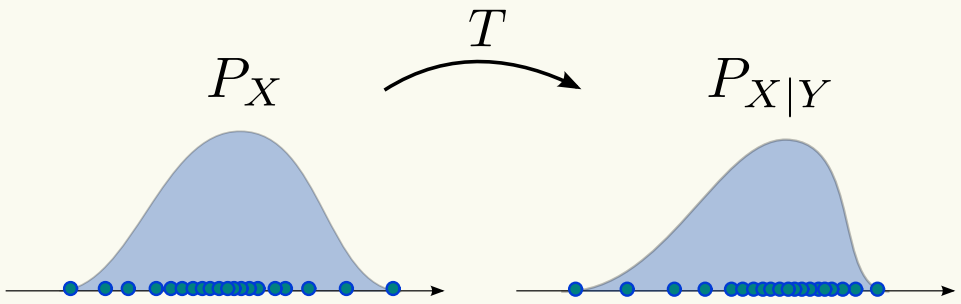
Example:

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

Questions: In a general setting,

- does the map exist?
- how to numerically find one?

Conditioning with transport maps



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

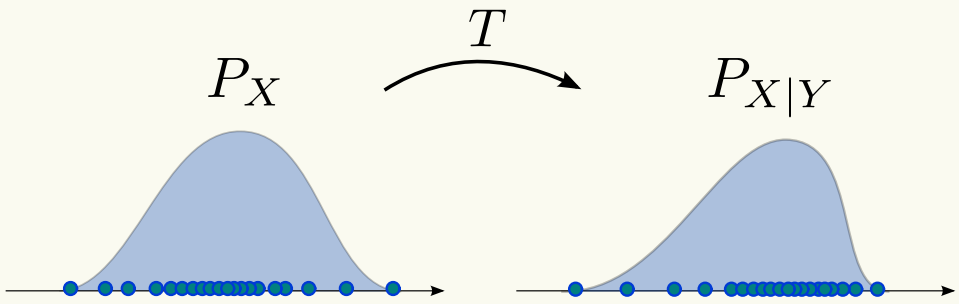
Example:

- 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?

Conditioning with transport maps



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

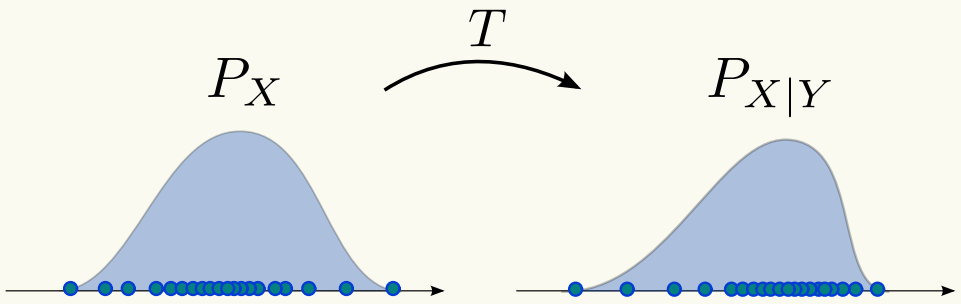
Example:

- 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?

Conditioning with transport maps



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

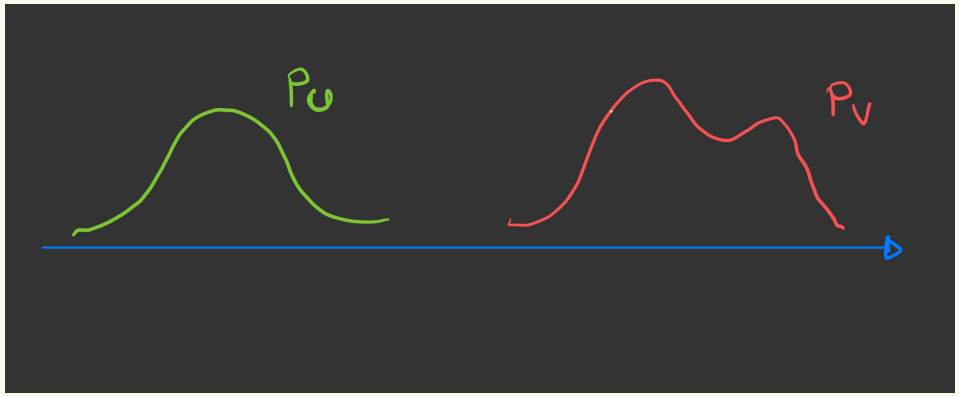
Example:

- 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?

Background on optimal transportation theory
Monge problem and Brenier's result



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

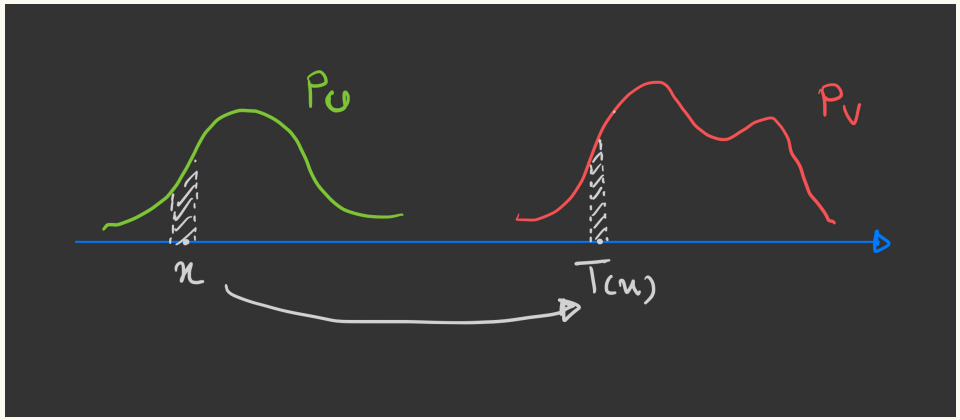
Questions:

- Does the optimal map exist? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? semi-dual Kantorovich problem

$$\inf_{T_{\#}P_U = P_V} \int \|T(x) - x\|^2 dP_U(x)$$

A. Vardhan Makkuva, A. Taghvaei, J. D. Lee, S. Oh, "Optimal transport mapping via input convex neural networks", ICML,. 2020

Background on optimal transportation theory
Monge problem and Brenier's result



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

Questions:

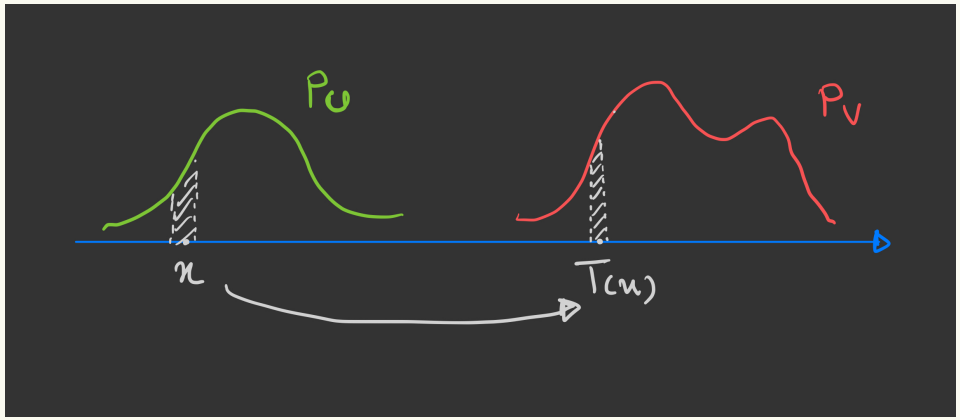
- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? semi-dual Kantorovich problem:

$$\max_{\psi} \left[\int \psi^c(y) dP_V(y) + \int \psi(x) dP_U(x) \right]$$

A. Vardhan Makkuva, A. Taghvaei, J. D. Lee, S. Oh, "Optimal transport mapping via input convex neural networks", ICML,. 2020

Background on optimal transportation theory

Monge problem and Brenier's result



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

Questions:

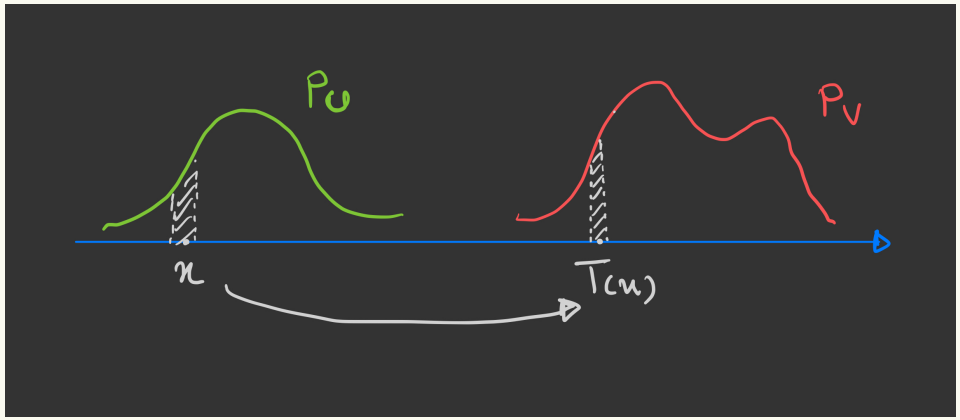
- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? semi-dual Kantorovich problem...

$$\max_{\psi} \left[\int \psi^c(y) dP_V(y) + \int \psi(x) dP_U(x) \right]$$

A. Vardhan Makkuva, A. Taghvaei, J. D. Lee, S. Oh, "Optimal transport mapping via input convex neural networks", ICML,. 2020

Background on optimal transportation theory

Monge problem and Brenier's result



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

Questions:

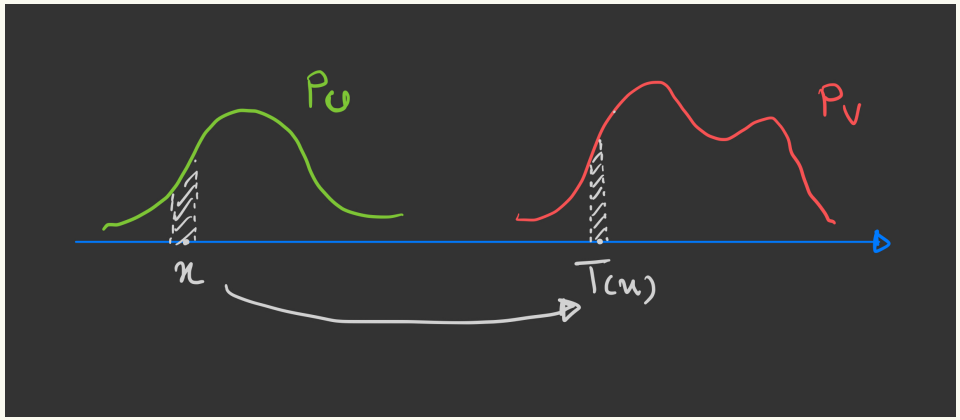
- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? semi-dual Kantorovich problem

$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 - f(T(U)) + f(V)\right]$$

A. Vardhan Makkuva, A. Taghvaei, J. D. Lee, S. Oh, "Optimal transport mapping via input convex neural networks", ICML,. 2020

Background on optimal transportation theory

Monge problem and Brenier's result



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

Questions:

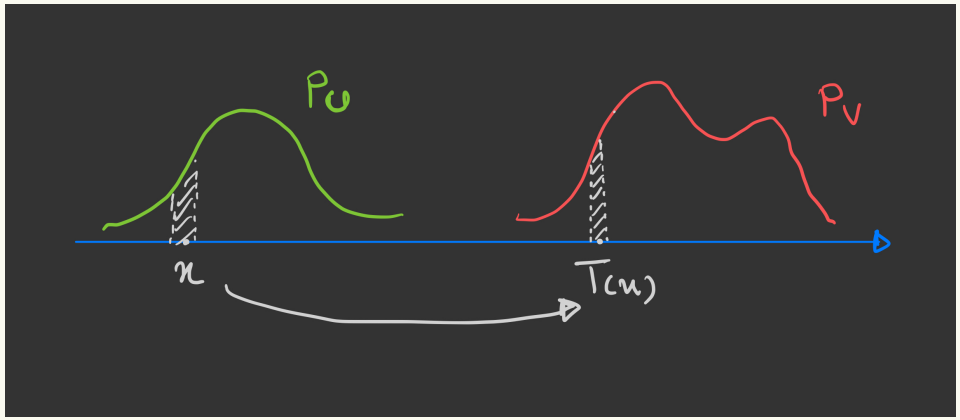
- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? semi-dual Kantorovich problem

$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 - f(T(U)) + f(V)\right]$$

A. Vardhan Makkuva, A. Taghvaei, J. D. Lee, S. Oh, "Optimal transport mapping via input convex neural networks", ICML,. 2020

Background on optimal transportation theory

Monge problem and Brenier's result



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

Questions:

- Does the optimal map exists? Yes, as long as P_U admits Lebesgue density
- How to numerically find it? semi-dual Kantorovich problem

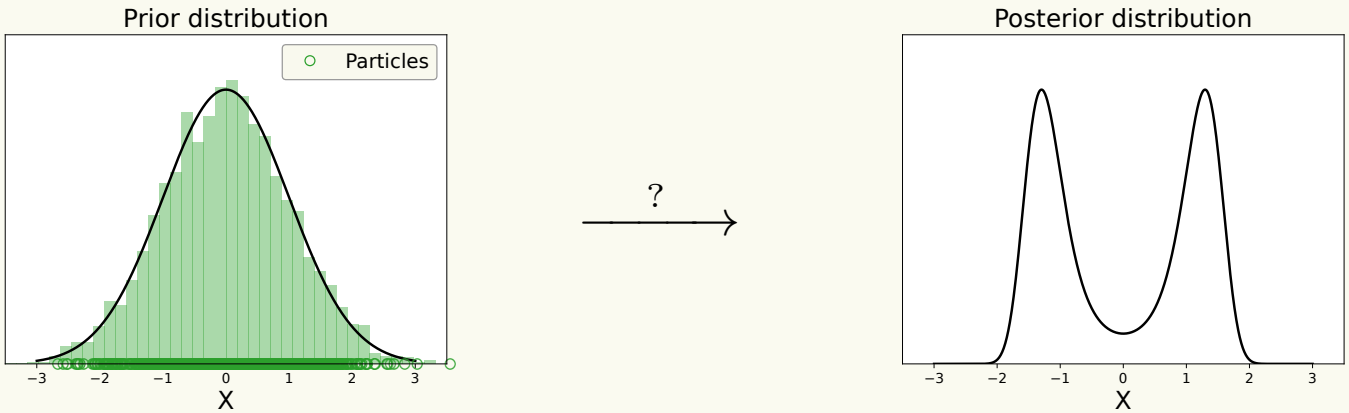
$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 - f(T(U)) + f(V)\right]$$

A. Vardhan Makkuva, A. Taghvaei, J. D. Lee, S. Oh, "Optimal transport mapping via input convex neural networks", ICML,. 2020

Conditioning with optimal transport map

Illustrative example

$\rightarrow$

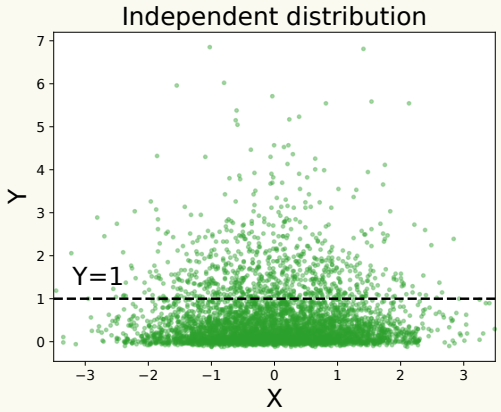


$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

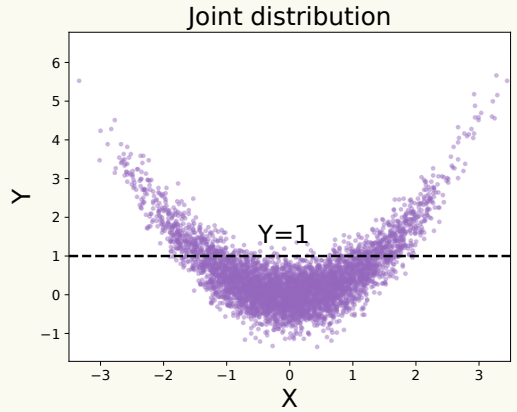
A. Taghvaei, B. Hosseini, "An optimal transport formulation of Bayes' law for nonlinear filtering algorithms", CDC, 2022
M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024

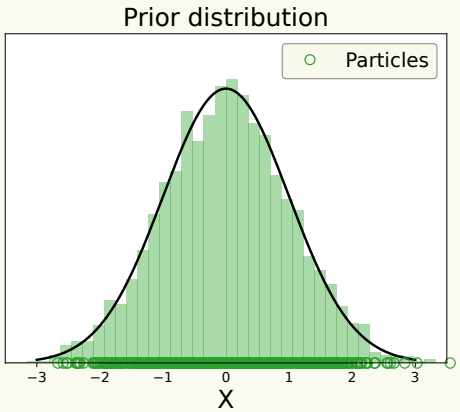
Conditioning with optimal transport map

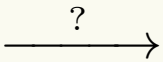
Illustrative example

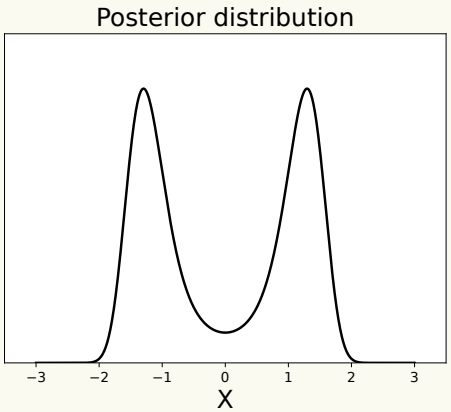












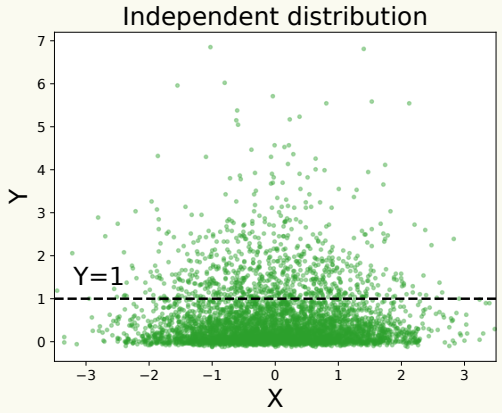
$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

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

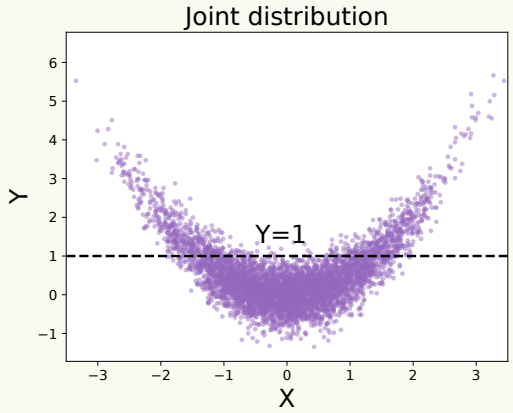
M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024

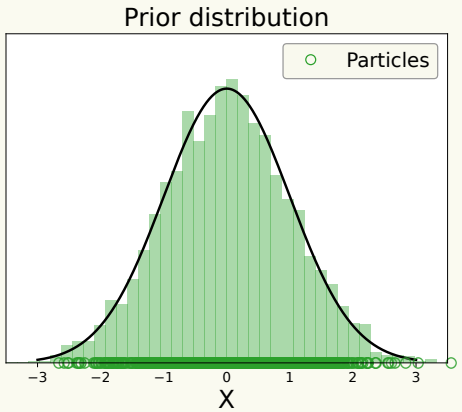
Conditioning with optimal transport map

Illustrative example

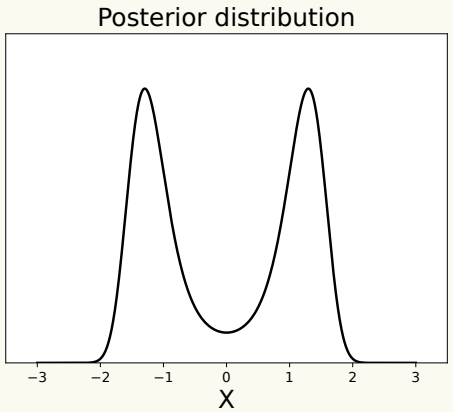


$$(T(X,Y),Y)$$





$$?$$

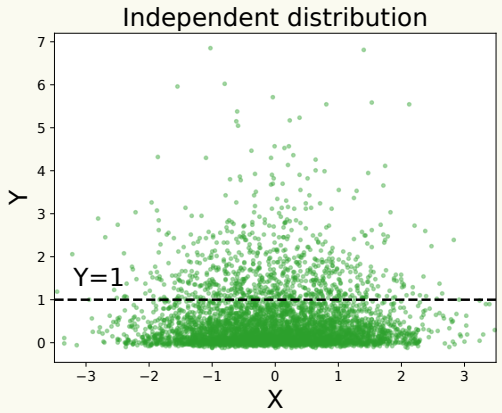


$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

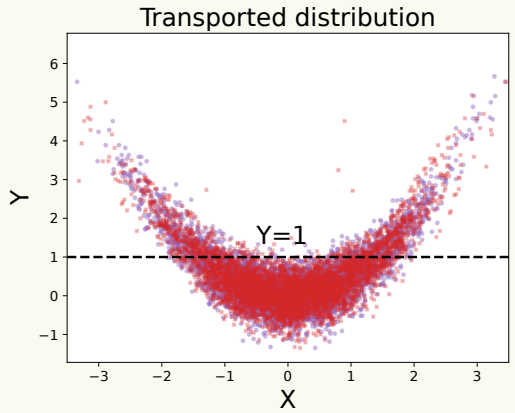
A. Taghvaei, B. Hosseini, "An optimal transport formulation of Bayes' law for nonlinear filtering algorithms", CDC, 2022
M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024

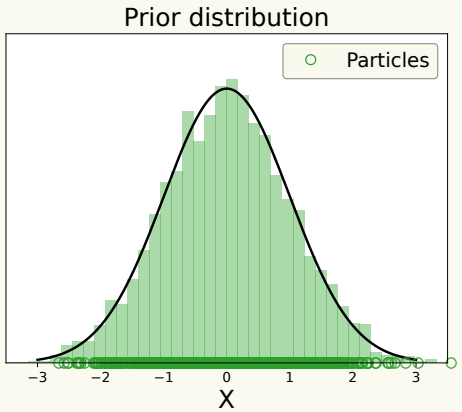
Conditioning with optimal transport map

Illustrative example

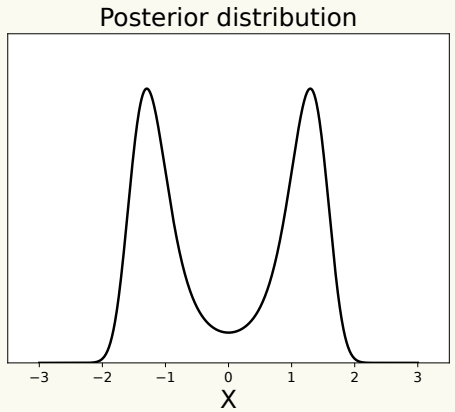


$$\xrightarrow{(T(X,Y),Y)}$$





$$\xrightarrow{?}$$

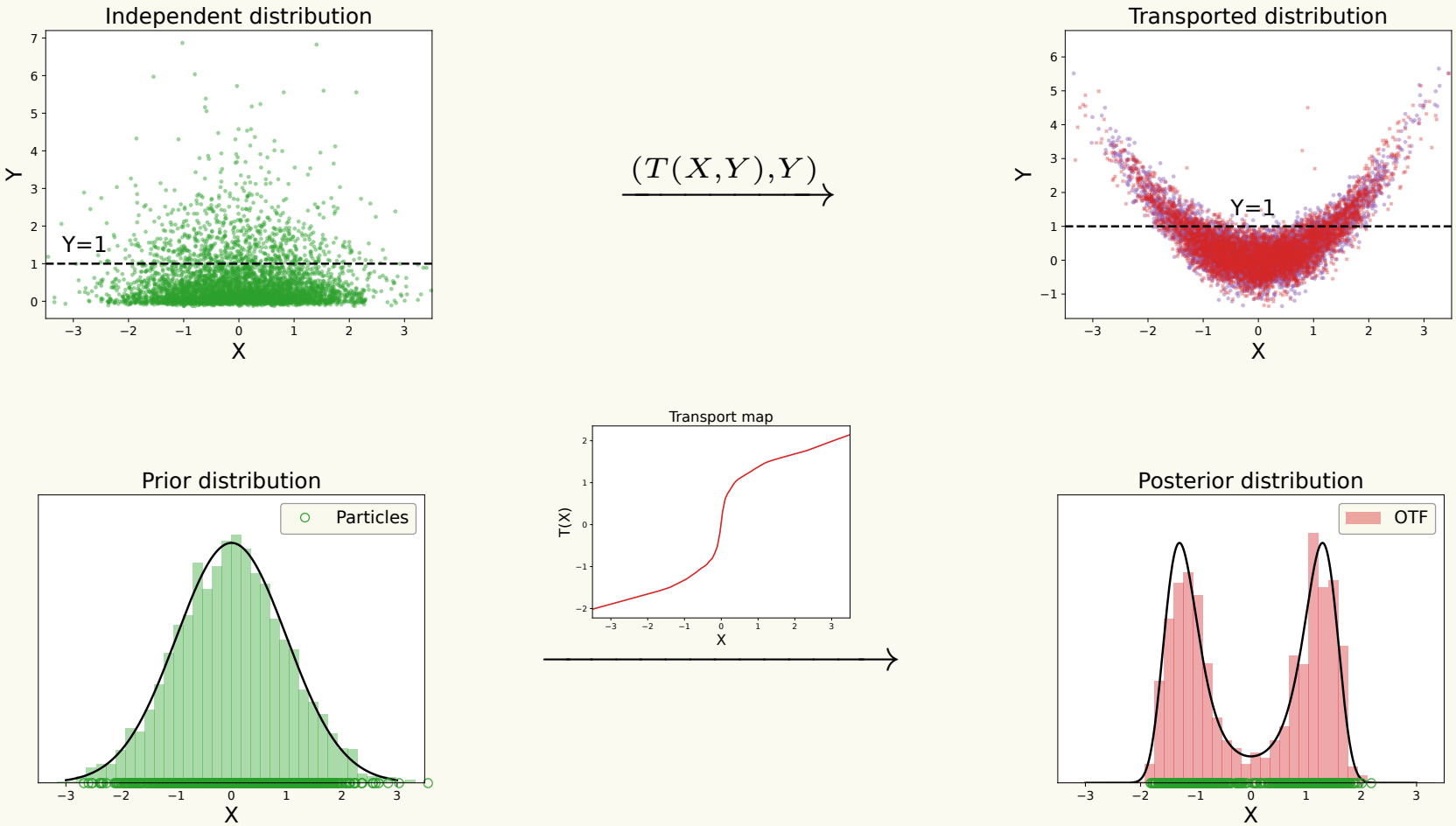


$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

A. Taghvaei, B. Hosseini, "An optimal transport formulation of Bayes' law for nonlinear filtering algorithms", CDC, 2022
M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024

Conditioning with optimal transport map

Illustrative example



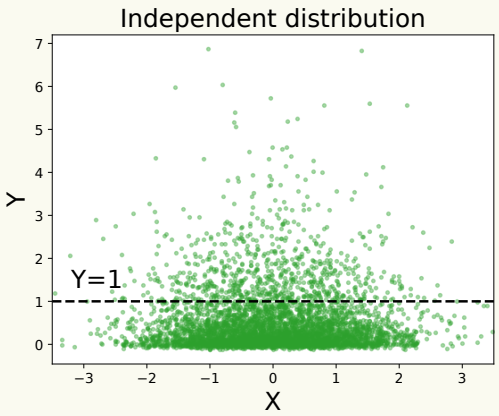
$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

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

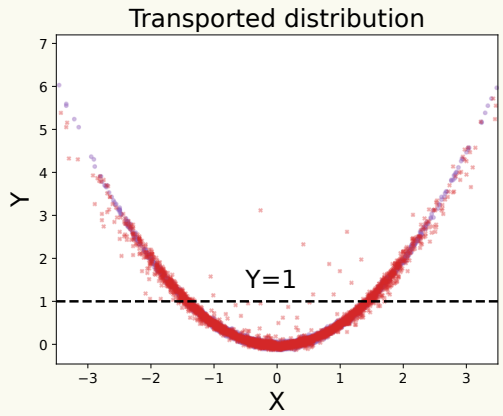
M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024

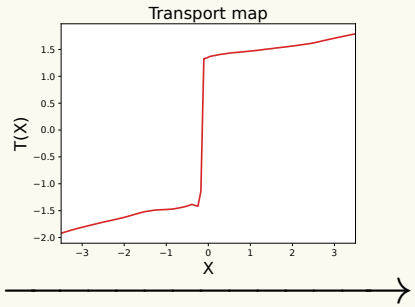
Conditioning with optimal transport map

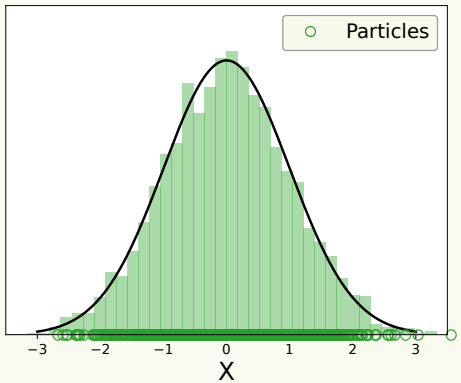
Illustrative example

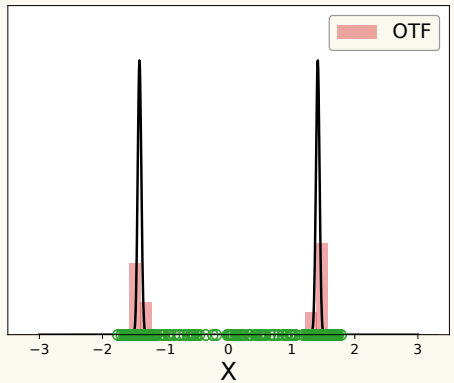


$$(T(X,Y),Y) \longrightarrow$$









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

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

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

Conditioning with optimal transport map

Variational formulation of Bayes' law

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

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

Conditional semi-dual Kantorovich formulation:

$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X}, Y) - \bar{X}\|^2 - f(T(\bar{X}, Y), Y) + f(X; Y)\right]$$

solved from samples · T, f represented by neural networks

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)

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338.

Conditioning with optimal transport map

Variational formulation of Bayes' law

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

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

Conditional semi-dual Kantorovich formulation:

$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X}, Y) - \bar{X}\|^2 - f(T(\bar{X}, Y), Y) + f(X; Y)\right]$$

solved from samples · T, f represented by neural networks

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)

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338.

Conditioning with optimal transport map

Variational formulation of Bayes' law

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

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

Conditional semi-dual Kantorovich formulation:

$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X}, Y) - \bar{X}\|^2 - f(T(\bar{X}, Y), Y) + f(X; Y)\right]$$

solved from samples · T, f represented by neural networks

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)

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338.

Conditioning with optimal transport map

Variational formulation of Bayes' law

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

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

Conditional semi-dual Kantorovich formulation:

$$\max_f \min_T \mathbb{E}\left[\frac{1}{2}\|T(\bar{X}, Y) - \bar{X}\|^2 - f(T(\bar{X}, Y), Y) + f(X; Y)\right]$$

solved from samples · T, f represented by neural networks

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)

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338.

Conditioning with optimal transport map
Theoretical analysis

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

(Conditional) Brenier's theorem

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

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

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

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

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338

Conditioning with optimal transport map
Theoretical analysis

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

(Conditional) Brenier's theorem

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

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

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

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

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338

Conditioning with optimal transport map
Theoretical analysis

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

(Conditional) Brenier's theorem

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

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

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

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

M. Al-Jarrah, N. Jin, B. Hosseini, A. Taghvaei, "Nonlinear Filtering with Brenier Optimal Transport Maps", ICML, 2024.
B. Hosseini, A. Hsu, A. Taghvaei, "Conditional Optimal Transport on Function Spaces", SIAM/ASA Journal on Uncertainty Quantification, 13(1), 304-338

Conditioning with optimal transport map

Statistical error analysis

$\phi(\cdot, y) := \frac{1}{2}\|\cdot\|^2 - f(\cdot, y)$ and $T = \nabla_x \phi$

$$\text{(population limit)} \quad \phi^\dagger := \underset{\phi \in \mathrm{CVX}_x}{\arg\min}\ \mathbb{E}\left[\phi(\overline{X}, Y) + \phi^\star(X, Y)\right]$$

$$\text{(empirical minimization)} \quad \hat{\phi} := \underset{\phi \in \mathcal{F}}{\arg\min}\ \frac{1}{N}\sum_{i=1}^{N} \phi(\overline{X}_i, Y_i) + \phi^\star(X_i, Y_i)$$

Theorem

Assume,

- **Boundedness**: $\sup_{\phi \in \mathcal{F}} \|\phi\|_{L^\infty} \le R < \infty$
- **Regularity**: For all $\phi \in \mathcal{F}$ and a.e. y, $\phi(\cdot, y)$ is α-strongly convex and β-smooth
- **Complexity control**: $\log \mathcal{N}(\delta, \mathcal{F}, \|\cdot\|_{L^\infty}) \le C_{\mathcal{F}} \delta^{-\gamma} \log(1 + \delta^{-1})$

Then,

$$\mathbb{E}\left[\|\nabla_x \hat{\phi} - \nabla_x \phi^\dagger\|_{L^2}^2\right] \le (\text{const.}) \left[\frac{\beta}{\alpha} \underbrace{\inf_{\phi \in \mathcal{F}} \|\nabla_x \phi - \nabla_x \phi^\dagger\|_{L^2}^2}_{\text{Approximation Bias}} + \underbrace{\sqrt{\frac{R^2 \beta^2 C_{\mathcal{F}}}{N}}}_{\text{Generalization error}}\right]$$

M. Al-Jarrah, B. Hosseini, N. Jin, M. Martino, A. Taghvaei, "Error Analysis of Triangular Optimal Transport Maps for Filtering", arXiv:2510.19283 (2025)

Conditioning with optimal transport map

Statistical error analysis

$\phi(\cdot, y) := \frac{1}{2}\|\cdot\|^2 - f(\cdot, y)$ and $T = \nabla_x \phi$

$$\text{(population limit)} \quad \phi^\dagger := \operatorname*{arg\,min}_{\phi \in \mathrm{CVX}_x} \mathbb{E}\left[\phi(\overline{X}, Y) + \phi^\star(X, Y)\right]$$

$$\text{(empirical minimization)} \quad \hat{\phi} := \operatorname*{arg\,min}_{\phi \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} \phi(\overline{X}_i, Y_i) + \phi^\star(X_i, Y_i)$$

Theorem

Assume,

- **Boundedness**: $\sup_{\phi \in \mathcal{F}} \|\phi\|_{L^\infty} \le R < \infty$
- **Regularity**: For all $\phi \in \mathcal{F}$ and a.e. y, $\phi(\cdot, y)$ is α-strongly convex and β-smooth
- **Complexity control**: $\log \mathcal{N}(\delta, \mathcal{F}, \|\cdot\|_{L^\infty}) \le C_{\mathcal{F}} \delta^{-\gamma} \log(1 + \delta^{-1})$

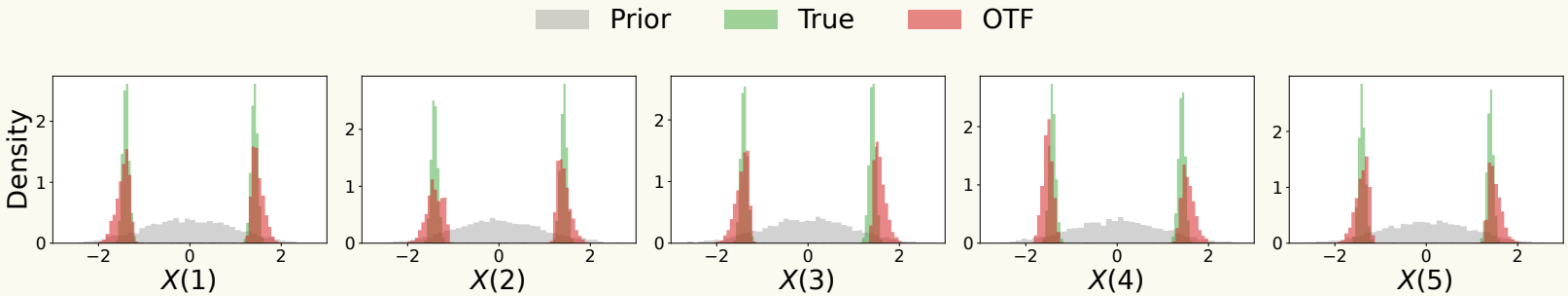
Then,

$$\mathbb{E}\left[\|\nabla_x \hat{\phi} - \nabla_x \phi^\dagger\|_{L^2}^2\right] \le \text{(const.)} \left[\frac{\beta}{\alpha} \underbrace{\inf_{\phi \in \mathcal{F}} \|\nabla_x \phi - \nabla_x \phi^\dagger\|_{L^2}^2}_{\text{Approximation Bias}} + \underbrace{\sqrt{\frac{R^2 \beta^2 C_{\mathcal{F}}}{N}}}_{\text{Generalization error}}\right]$$

M. Al-Jarrah, B. Hosseini, N. Jin, M. Martino, A. Taghvaei, "Error Analysis of Triangular Optimal Transport Maps for Filtering", arXiv:2510.19283 (2025)

Computational properties

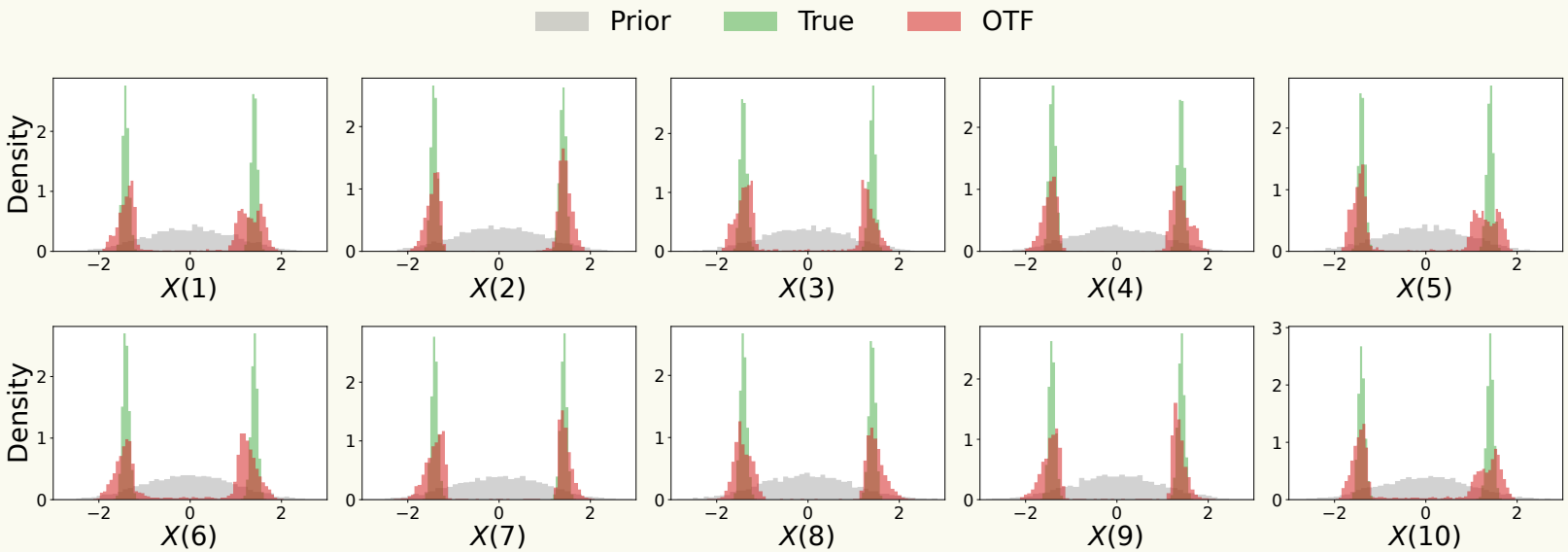
Scalability with the dimension: $X \sim \mathcal{N}(0, I_d),\ Y = X \odot X + W$



$$\text{dim} = 5 \quad \rightarrow \quad \text{number of modes} = 32$$

Computational properties

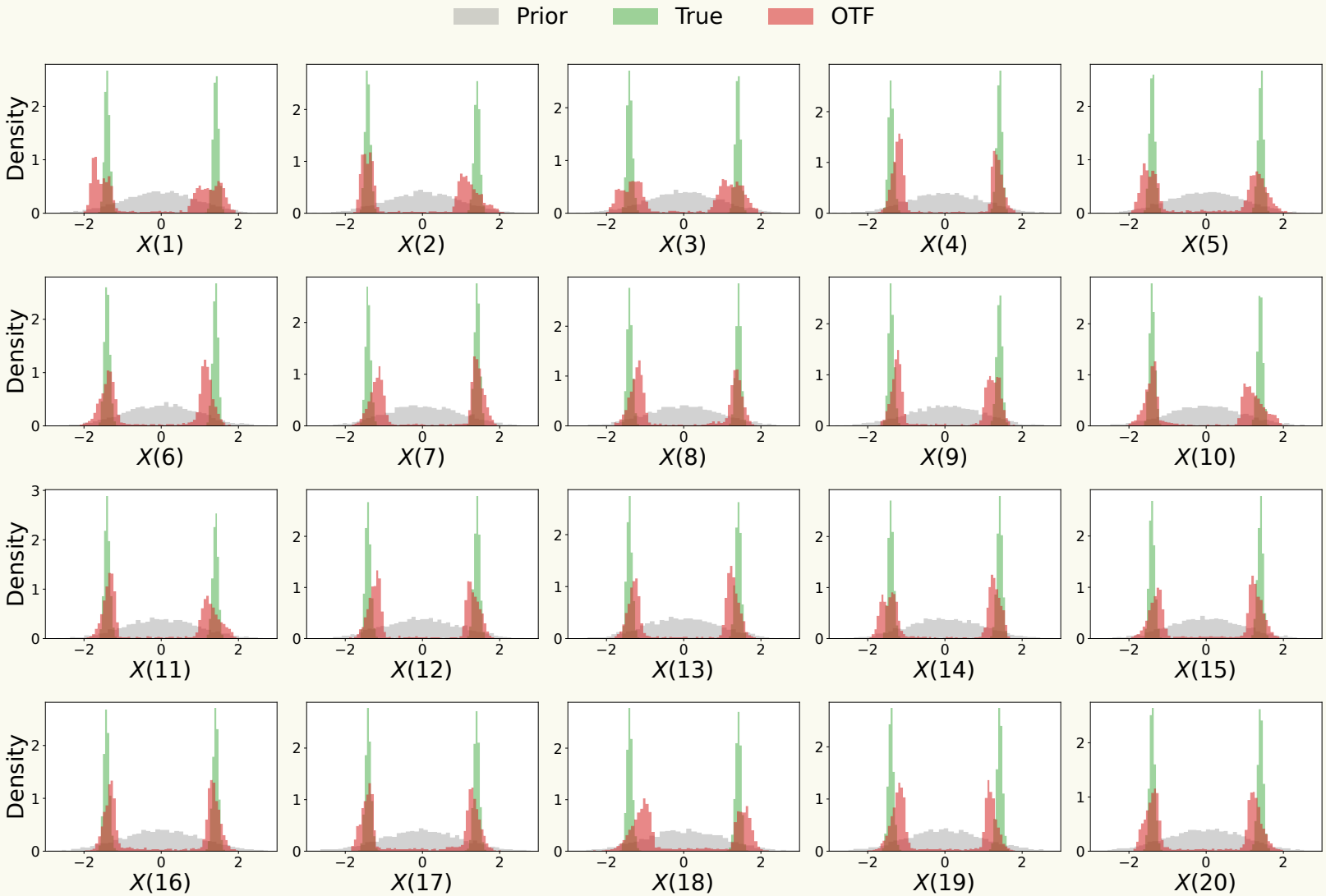
Scalability with the dimension: $X \sim \mathcal{N}(0, I_d),\ Y = X \odot X + W$



$$\text{dim} = 10 \quad \rightarrow \quad \text{number of modes} = 1052$$

Computational properties

Scalability with the dimension: $X \sim \mathcal{N}(0, I_d),\ Y = X \odot X + W$



$$\text{dim} = 20 \quad \rightarrow \quad \text{number of modes} \approx 1000000$$

Outline

- Foundations
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Outline

- Foundations
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Nonlinear filtering problem

Model:

$$\begin{aligned} \text{dynamic:} \quad & X_t = a(X_{t-1}, V_t) \\ \text{observation:} \quad & Y_t = h(X_t, W_t) \end{aligned}$$

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

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

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

Nonlinear filtering: generate samples from the posterior π_t for all t.

$$\{X_t^i\}_{i=1}^N \sim \pi_t$$

Nonlinear filtering problem

Model:

$$\begin{aligned} \text{dynamic:} \quad & X_t = a(X_{t-1}, V_t) \\ \text{observation:} \quad & Y_t = h(X_t, W_t) \end{aligned}$$

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

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

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

Nonlinear filtering: generate samples from the posterior π_t for all t.

$$\{X_t^i\}_{i=1}^N \sim \pi_t$$

Nonlinear filtering problem

Model:

$$\begin{aligned} \text{dynamic:} \quad & X_t = a(X_{t-1}, V_t) \\ \text{observation:} \quad & Y_t = h(X_t, W_t) \end{aligned}$$

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

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

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

Nonlinear filtering: generate samples from the posterior π_t for all t.

$$\{X_t^i\}_{i=1}^N \sim \pi_t$$

Nonlinear filtering problem

Model:

$$\begin{aligned} \text{dynamic:} \quad & X_t = a(X_{t-1}, V_t) \\ \text{observation:} \quad & Y_t = h(X_t, W_t) \end{aligned}$$

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

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

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

Nonlinear filtering: generate samples from the posterior π_t for all t.

$$\{X_t^i\}_{i=1}^N \sim \pi_t$$

Optimal Transport Filter (OTF)
Design

Two important operations:

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

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

Update law for the exact posterior:

$$\text{propagation:} \quad \pi_{t|t-1} = \mathcal{A}\pi_{t-1}$$

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

OT filter:

$$\text{propagation:} \quad X^i_{t|t-1} = a(X^i_{t-1}, V^i_t), \quad Y^i_{t|t-1} = h(X^i_{t|t-1}, W^i_t)$$

$$\text{conditioning:} \quad X^i_t = T_t(X^i_{t|t-1}, Y_t)$$

$$\text{where} \quad T_t \leftarrow \max_f \min_T J(f, T, \{X^i_{t|t-1}, Y^i_{t|t-1}\}_{i=1}^N)$$

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

Optimal Transport Filter (OTF)
Design

Two important operations:

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

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

Update law for the exact posterior:

$$\text{propagation:} \quad \pi_{t|t-1} = \mathcal{A}\pi_{t-1}$$

$$\text{conditioning:} \quad \pi_t = \mathcal{B}_{Y_t}\left(\pi_{t|t-1}\right)$$

OT filter:

$$\text{propagation:} \quad X^i_{t|t-1} = a(X^i_{t-1}, V^i_t), \quad Y^i_{t|t-1} = h(X^i_{t|t-1}, W^i_t)$$

$$\text{conditioning:} \quad X^i_t = T_t(X^i_{t|t-1}, Y_t)$$

$$\text{where} \quad T_t \leftarrow \max_f \min_T J(f, T, \{X^i_{t|t-1}, Y^i_{t|t-1}\}_{i=1}^N)$$

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

Optimal Transport Filter (OTF)
Design

Two important operations:

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

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

Update law for the exact posterior:

$$\text{propagation:} \quad \pi_{t|t-1} = \mathcal{A}\pi_{t-1}$$

$$\text{conditioning:} \quad \pi_t = \mathcal{B}_{Y_t}\left(\pi_{t|t-1}\right)$$

OT filter:

$$\text{propagation:} \quad X^i_{t|t-1} = a(X^i_{t-1}, V^i_t), \quad Y^i_{t|t-1} = h(X^i_{t|t-1}, W^i_t)$$

$$\text{conditioning:} \quad X^i_t = T_t(X^i_{t|t-1}, Y_t)$$

$$\text{where} \quad T_t \leftarrow \max_f \min_T J(f, T, \{X^i_{t|t-1}, Y^i_{t|t-1}\}_{i=1}^N)$$

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: $X_{t|t-1}^i = a(X_{t-1}^i, V_t^i)$ and $Y_{t|t-1}^i = h(X_{t|t-1}^i, W_t^i)$
- 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

Algorithm

Initialize:

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

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

- propagation: $X_{t|t-1}^i = a(X_{t-1}^i, V_t^i)$ and $Y_{t|t-1}^i = h(X_{t|t-1}^i, W_t^i)$
- 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:

- Increase for a stable learning dynamic and observe overfitting
- In practice, only a few iterations of the optimization is performed

Algorithm

Initialize:

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

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

- propagation: $X_{t|t-1}^i = a(X_{t-1}^i, V_t^i)$ and $Y_{t|t-1}^i = h(X_{t|t-1}^i, W_t^i)$
- 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 a resampling step
- In practice, only a few iterations of the optimization are performed

Algorithm

Initialize:

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

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

- propagation: $X_{t|t-1}^i = a(X_{t-1}^i, V_t^i)$ and $Y_{t|t-1}^i = h(X_{t|t-1}^i, W_t^i)$
- 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:

- Increase for every filter iteration the cycle count and observe the results
- In practice, only a few iterations of the optimization are performed

Algorithm

Initialize:

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

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

- propagation: $X_{t|t-1}^i = a(X_{t-1}^i, V_t^i)$ and $Y_{t|t-1}^i = h(X_{t|t-1}^i, W_t^i)$
- 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

Algorithm

Initialize:

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

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

- propagation: $X_{t|t-1}^i = a(X_{t-1}^i, V_t^i)$ and $Y_{t|t-1}^i = h(X_{t|t-1}^i, W_t^i)$
- 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

Optimal Transport Filter
Numerical example

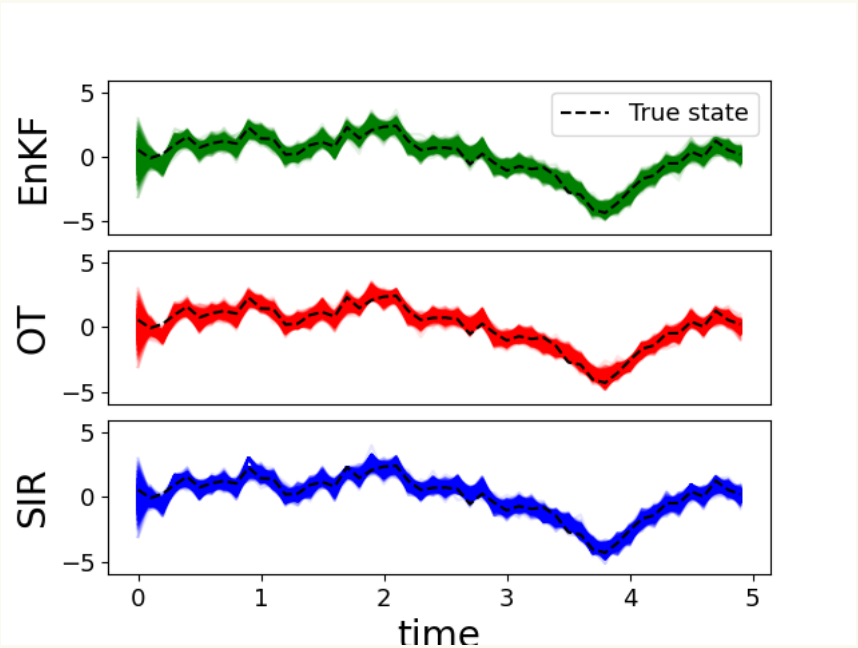
$$X_t = (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0,1),$$
$$Y_t = X_t + \sigma_W W_t,$$

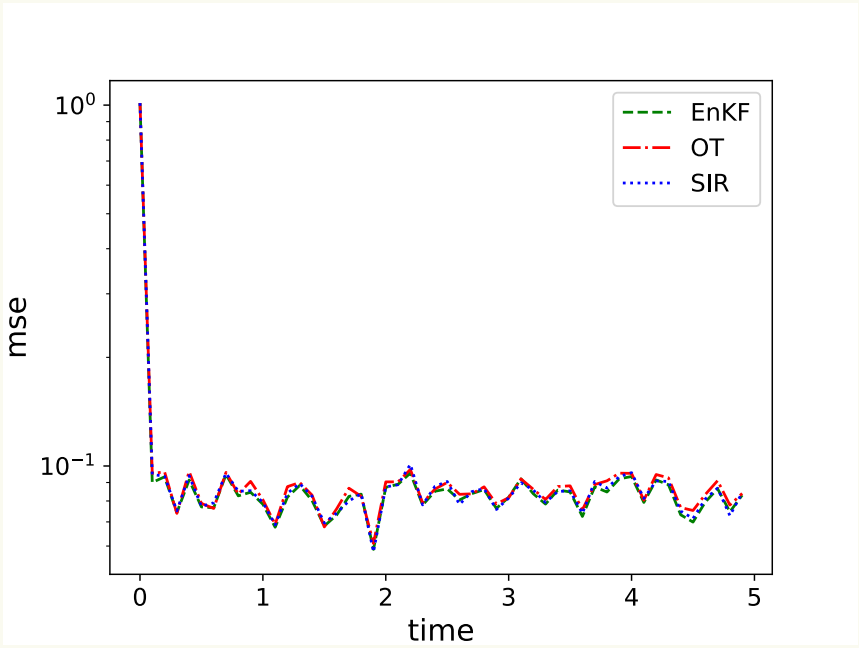
- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)

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

Optimal Transport Filter
Numerical example

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



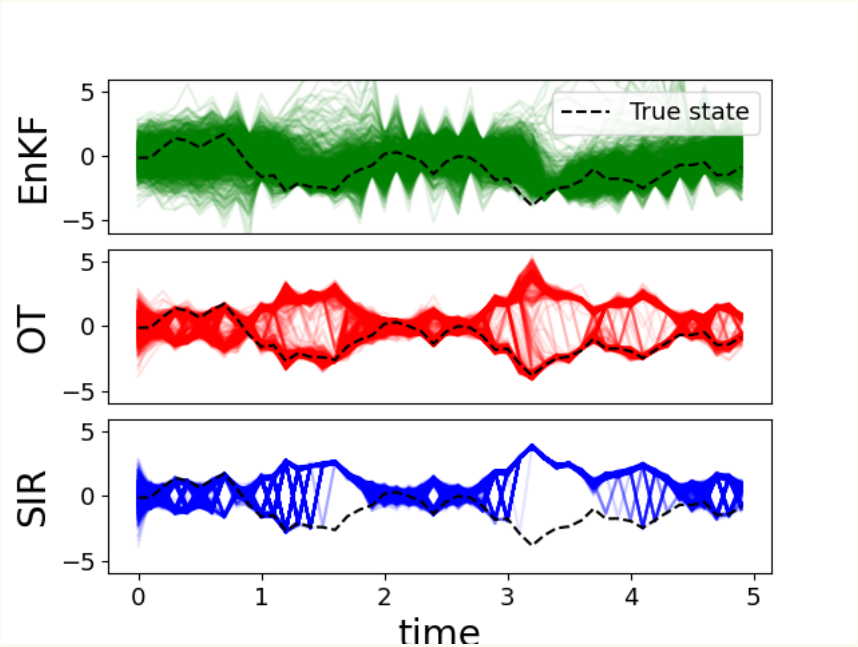


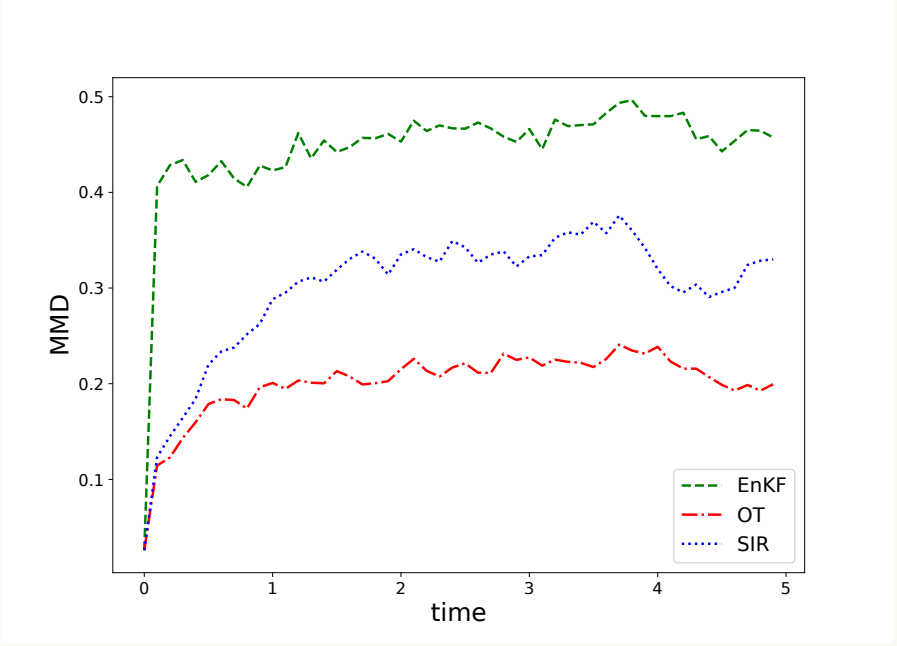
- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)

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

Optimal Transport Filter
Numerical example

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



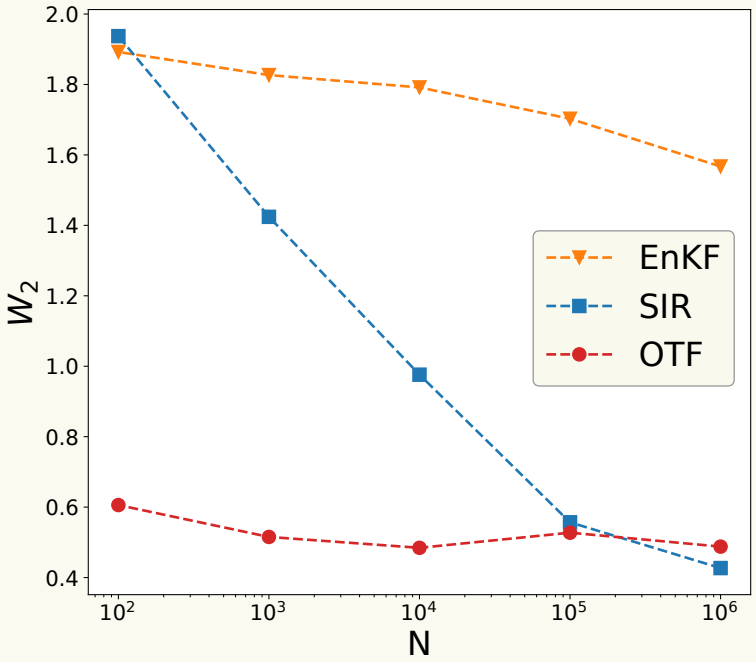


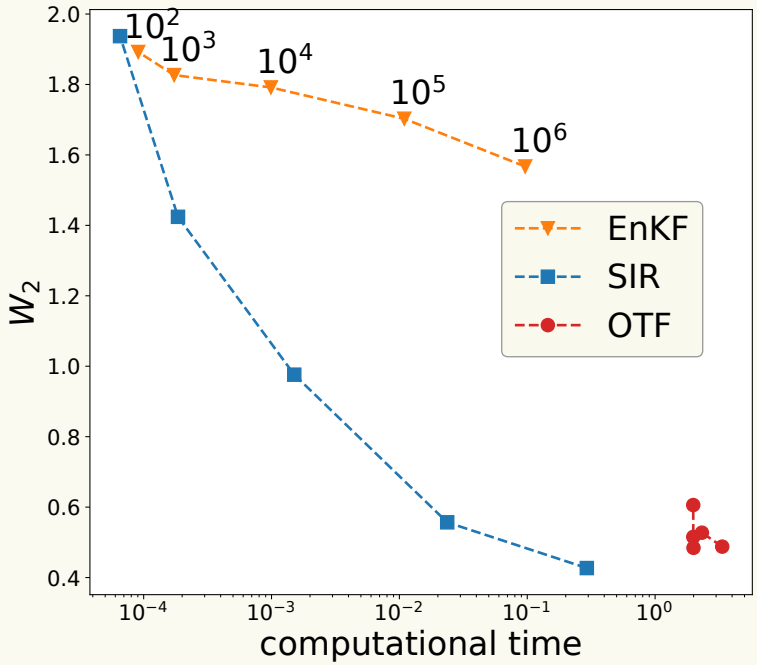
- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)

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

Computational properties

Sample and time complexity, speed-up with amortization (A-OTF)



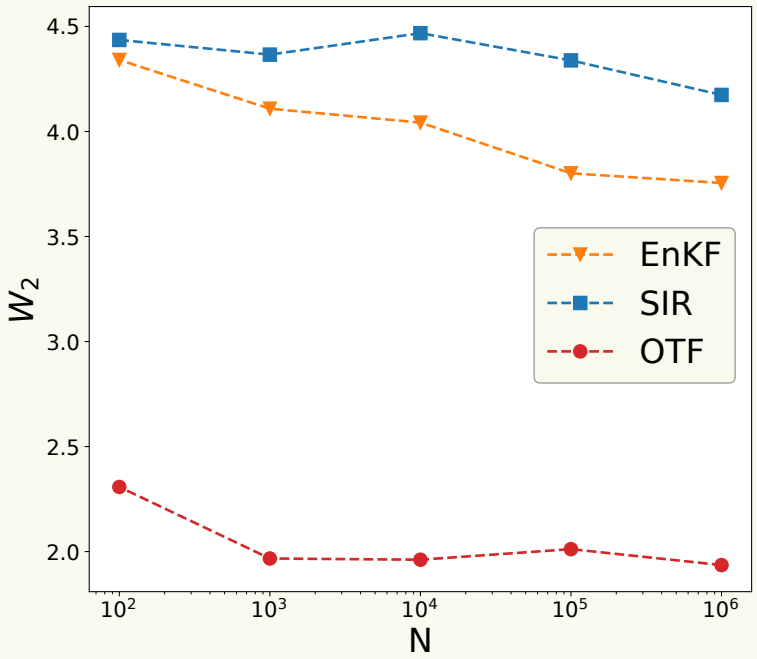


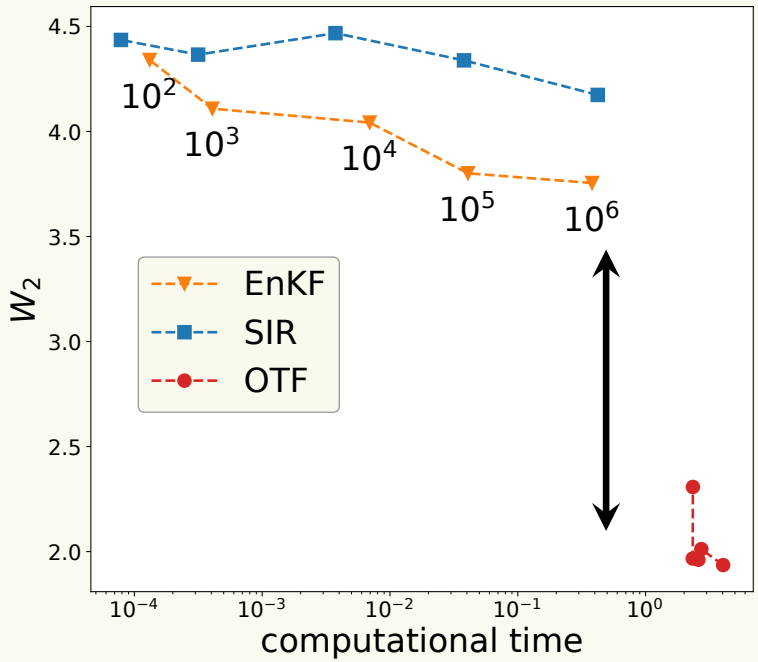
- dim $= 2$: OTF outperforms SIR when N is small
- dim $= 8$: SIR hits the memory limit
- A-OTF speeds up OTF while maintaining accuracy

M. Al-Jarrah, B. Hosseini, A. Taghvaei, "Fast filtering of non-Gaussian models using amortized optimal transport maps." L-CSS, pp. 1670-1675, 2025

Computational properties

Sample and time complexity, speed-up with amortization (A-OTF)



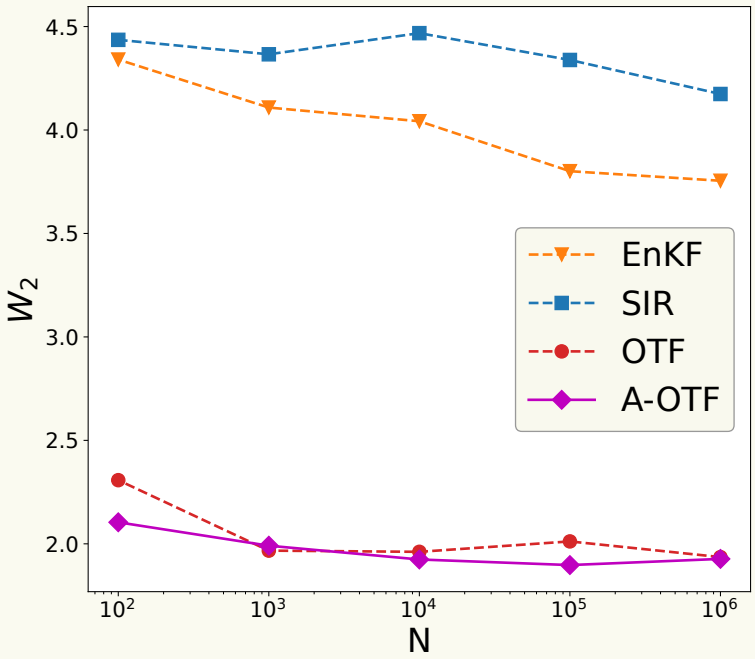


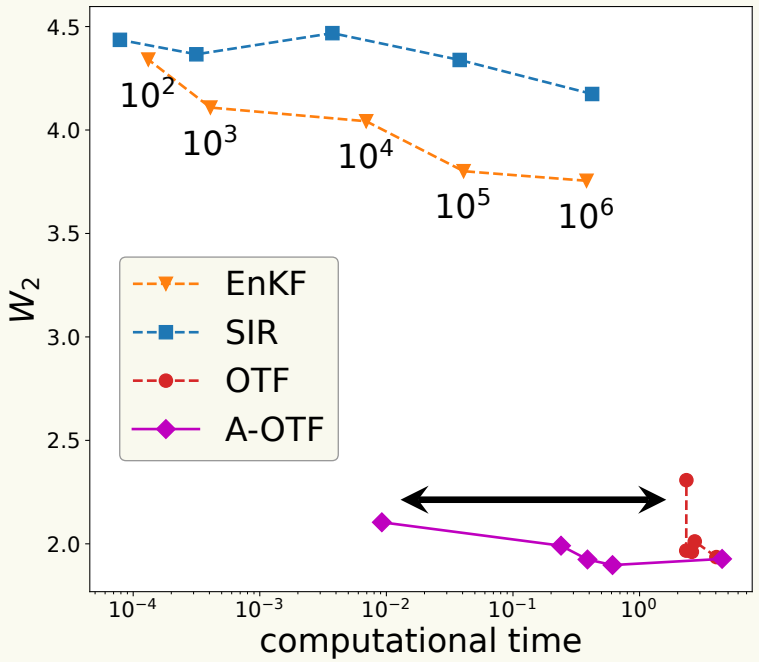
- dim $= 2$: OTF outperforms SIR when N is small
- dim $= 8$: SIR hits the memory limit
- A-OTF speeds up OTF while maintaining accuracy

M. Al-Jarrah, B. Hosseini, A. Taghvaei, "Fast filtering of non-Gaussian models using amortized optimal transport maps." L-CSS, pp. 1670-1675, 2025

Computational properties

Sample and time complexity, speed-up with amortization (A-OTF)





- dim $= 2$: OTF outperforms SIR when N is small
- dim $= 8$: SIR hits the memory limit
- A-OTF speeds up OTF while maintaining accuracy

M. Al-Jarrah, B. Hosseini, A. Taghvaei, "Fast filtering of non-Gaussian models using amortized optimal transport maps." L-CSS, pp. 1670-1675, 2025

Performance on benchmarks

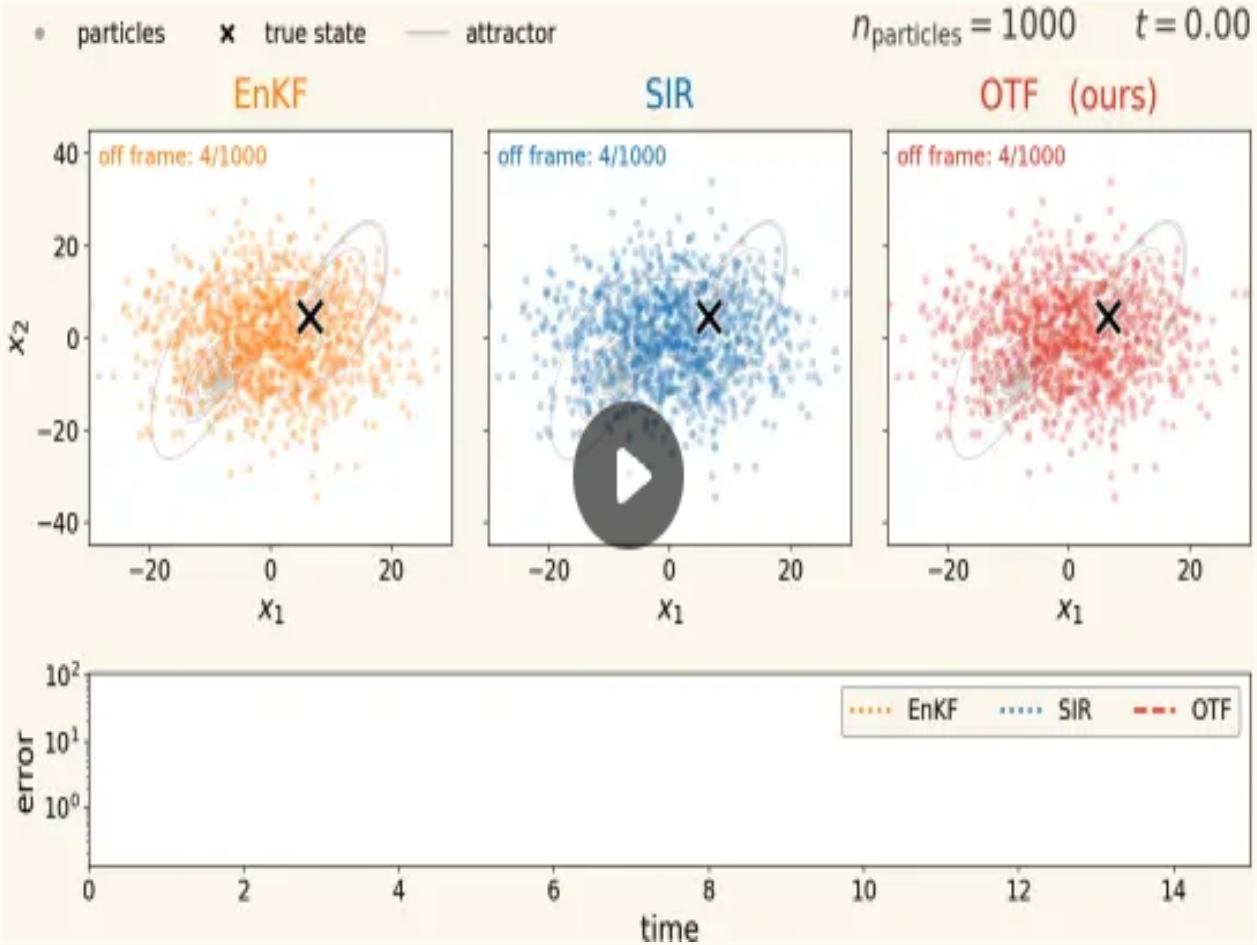
Lorentz 63

$$\begin{bmatrix} \dot{X}_t(1) \\ \dot{X}_t(2) \\ \dot{X}_t(3) \end{bmatrix} = \begin{bmatrix} \sigma(X_t(2) - X_t(1)) \\ X_t(1)(\rho - X_t(3)) - X_t(2) \\ X_t(1)X_t(2) - \beta X_t(3) \end{bmatrix}$$

$$Y_t = X_t(3) + \sigma W_t, \quad \sigma^2 = 1$$

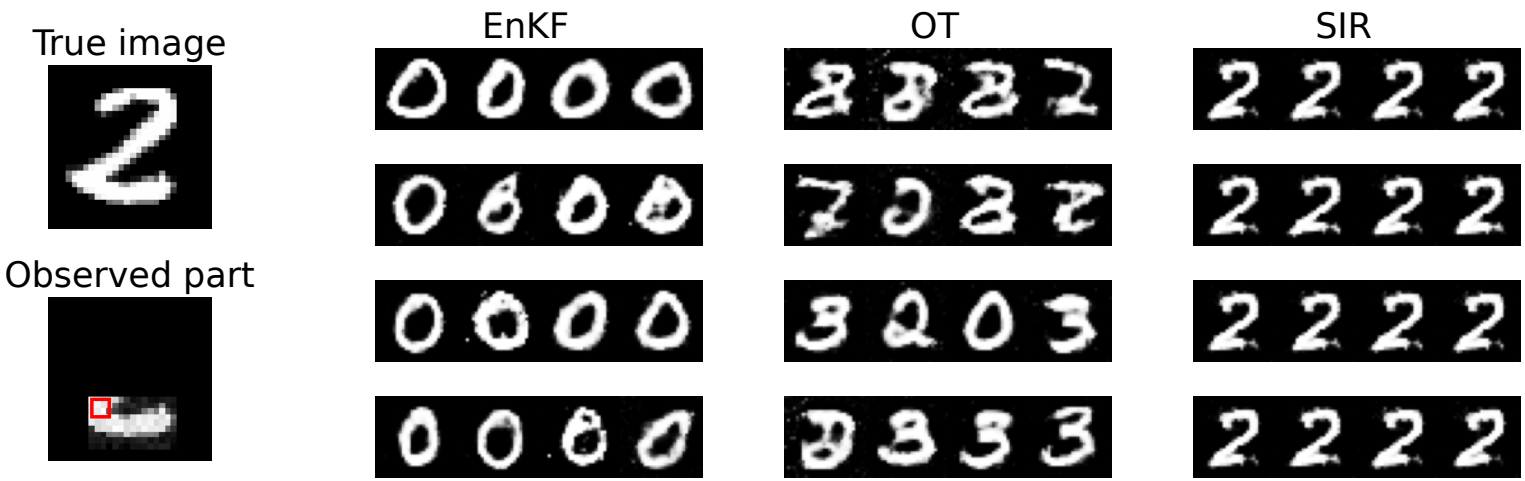
Performance on benchmarks

Lorentz 63



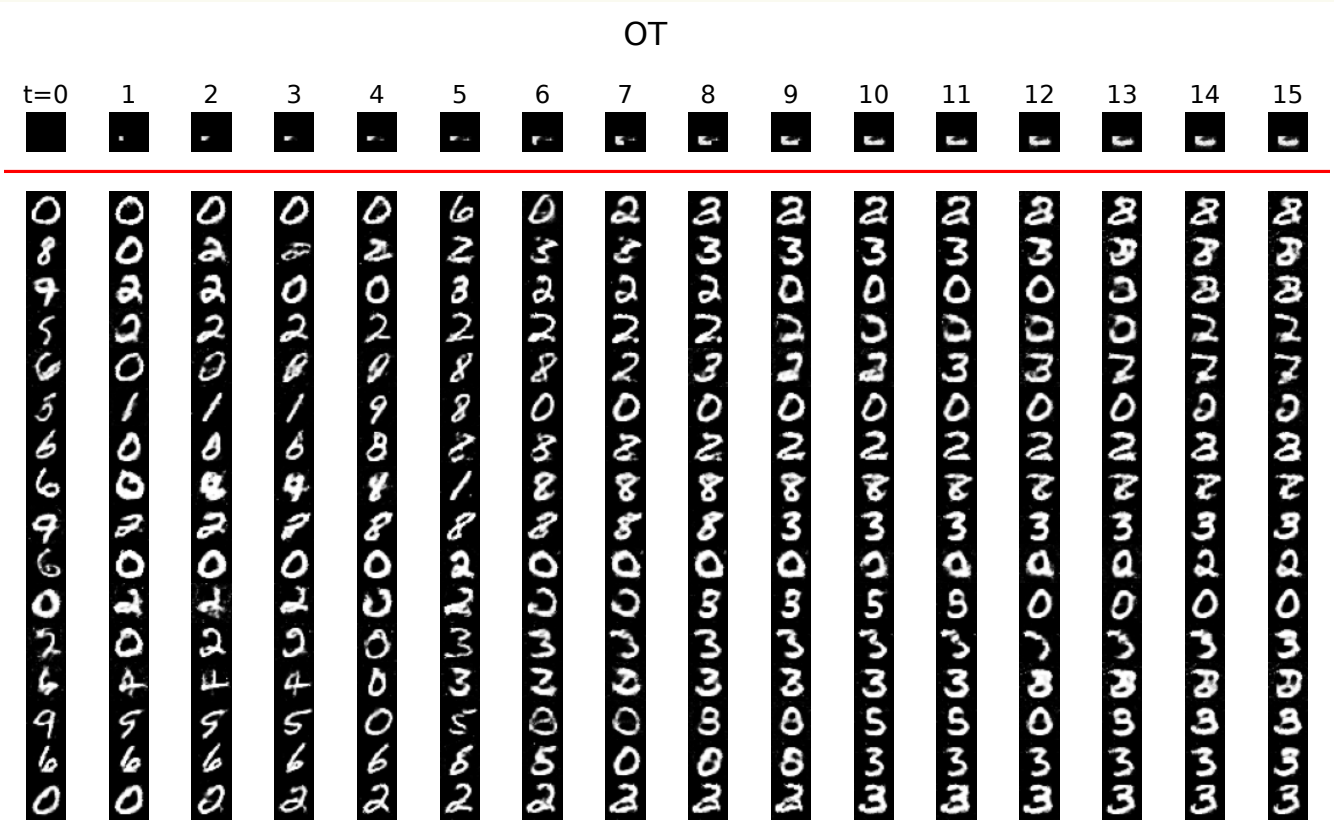
Numerical example: Image in-painting

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



Numerical example: Image in-painting

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



Extension to Riemannian manifolds

McCann's result

- Assume $X \in \mathcal{M}$ with metric g and geodesic distance d_g
- Replace the Euclidean distance with the geodesic distance
- Replace $T(x,y)$ with $\exp_x(U(x,y))$ where $U(x,y) \in T_x M$

$$\max_{f:\mathcal{M}\to\mathbb{R}} \min_{U:\mathcal{M}\to T\mathcal{M}} \mathbb{E}\left[\frac{1}{2} d_g(\exp_{\bar{X}}(U(\bar{X},Y)),\bar{X})^2 - f(\exp_{\bar{X}}(U(\bar{X},Y)),Y) + f(X;Y)\right]$$

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

Extension to Riemannian manifolds

McCann's result

- Assume $X \in \mathcal{M}$ with metric g and geodesic distance d_g
- Replace the Euclidean distance with the geodesic distance
- Replace $T(x,y)$ with $\exp_x(U(x,y))$ where $U(x,y) \in T_x M$

$$\max_{f:\mathcal{M}\to\mathbb{R}} \min_{U:\mathcal{M}\to T\mathcal{M}} \mathbb{E}\left[\frac{1}{2} d_g(\exp_{\bar{X}}(U(\bar{X},Y)),\bar{X})^2 - f(\exp_{\bar{X}}(U(\bar{X},Y)),Y) + f(X;Y)\right]$$

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

Extension to Riemannian manifolds

McCann's result

- Assume $X \in \mathcal{M}$ with metric g and geodesic distance d_g
- Replace the Euclidean distance with the geodesic distance
- Replace $T(x,y)$ with $\exp_x(U(x,y))$ where $U(x,y) \in T_x M$

$$\max_{f:\mathcal{M}\to\mathbb{R}} \min_{U:\mathcal{M}\to T\mathcal{M}} \mathbb{E}\left[\frac{1}{2} d_g(\exp_{\bar{X}}(U(\bar{X},Y)),\bar{X})^2 - f(\exp_{\bar{X}}(U(\bar{X},Y)),Y) + f(X;Y)\right]$$

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

Extension to Riemannian manifolds

McCann's result

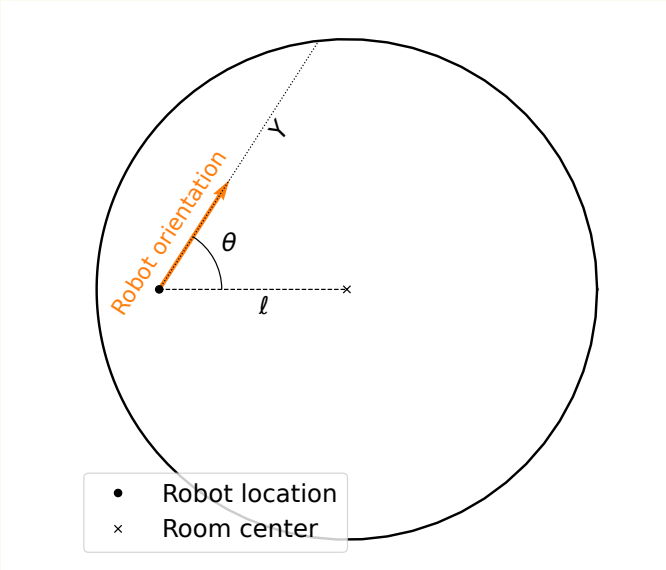
- Assume $X \in \mathcal{M}$ with metric g and geodesic distance d_g
- Replace the Euclidean distance with the geodesic distance
- Replace $T(x,y)$ with $\exp_x(U(x,y))$ where $U(x,y) \in T_x M$

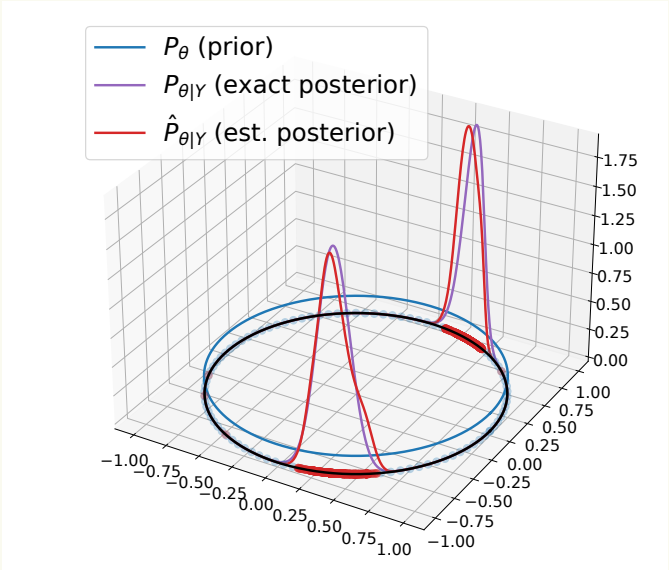
$$\max_{f:\mathcal{M}\to\mathbb{R}} \min_{U:\mathcal{M}\to T\mathcal{M}} \mathbb{E}\left[\frac{1}{2} d_g(\exp_{\bar{X}}(U(\bar{X},Y)),\bar{X})^2 - f(\exp_{\bar{X}}(U(\bar{X},Y)),Y) + f(X;Y)\right]$$

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

Numerical examples

- $\theta \in S^1$ is robot's orientation and Y is noisy measurement of distance to the wall
- $(x, \theta) \in SE(2)$ is robot's position and orientation.

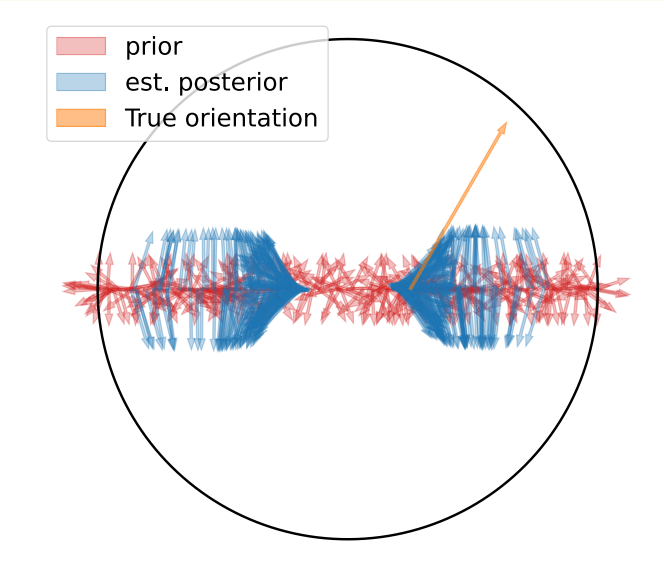


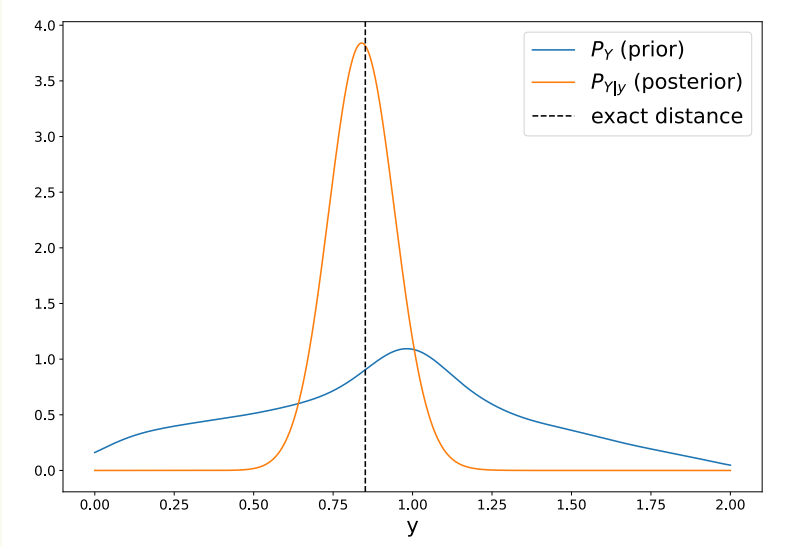


D. Grange, M. Al-Jarrah, R. Baptista, A. Taghvaei, T. Georgiou, S. Phillips, A. Tannenbaum, Computational optimal transport and filtering on Riemannian manifolds, IEEE Control Systems Letters, 2023

Numerical examples

- $\theta \in S^1$ is robot's orientation and Y is noisy measurement of distance to the wall
- $(x, \theta) \in SE(2)$ is robot's position and orientation.





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

Outline

- Foundations
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Outline

- Foundations
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Data-driven setting

Problem setup:

$$X_t = a(X_{t-1}, V_t), \quad X_0 \sim \pi_0$$
$$Y_t = h(X_t, W_t)$$

- 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 $\{Y_t, \ldots, Y_1\}$

Data-driven setting

Problem setup:

$$X_t = a(X_{t-1}, V_t), \quad X_0 \sim \pi_0$$
$$Y_t = h(X_t, W_t)$$

- 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 $\{Y_t, \ldots, Y_1\}$

Data-driven setting
Solution approach

- Exact posterior:

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

- Step 1: Truncated posterior

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

- Step 2: OT representation

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

- Step 3: Stationary assumption

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

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

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting
Solution approach

- Exact posterior:

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

- Step 1: Truncated posterior

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

- Step 2: OT representation

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

- Step 3: Stationary assumption

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

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

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting
Solution approach

- Exact posterior:

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

- Step 1: Truncated posterior

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

- Step 2: OT representation

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

- Step 3: Stationary assumption

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

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

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting
Solution approach

- Exact posterior:

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

- Step 1: Truncated posterior

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

- Step 2: OT representation

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

- Step 3: Stationary assumption

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

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

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting
Solution approach

- Exact posterior:

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

- Step 1: Truncated posterior

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

- Step 2: OT representation

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

- Step 3: Stationary assumption

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

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

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Data-driven setting

Error analysis

Assume

- The exact filter is exponentially stable
- The process (X_t, Y_t) is stationary
- μ is equal to the stationary distribution of X_t and $M := \sup_t d(\pi_t, \mu) < \infty$
- (f, T) is a possibly non-optimal pair with max-min gap $\epsilon(f, T)$
- The function $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \dots, y_1)$ is α-strongly convex for all $(y_w, \dots, y_1)$.

Then,

$$d(T(\cdot, Y_t, \dots, Y_{t-w})\#\mu, \pi_t) \leq CM(1-\lambda)^w + \sqrt{\frac{4}{\alpha}\epsilon(f, T)}$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Numerical example

Model:

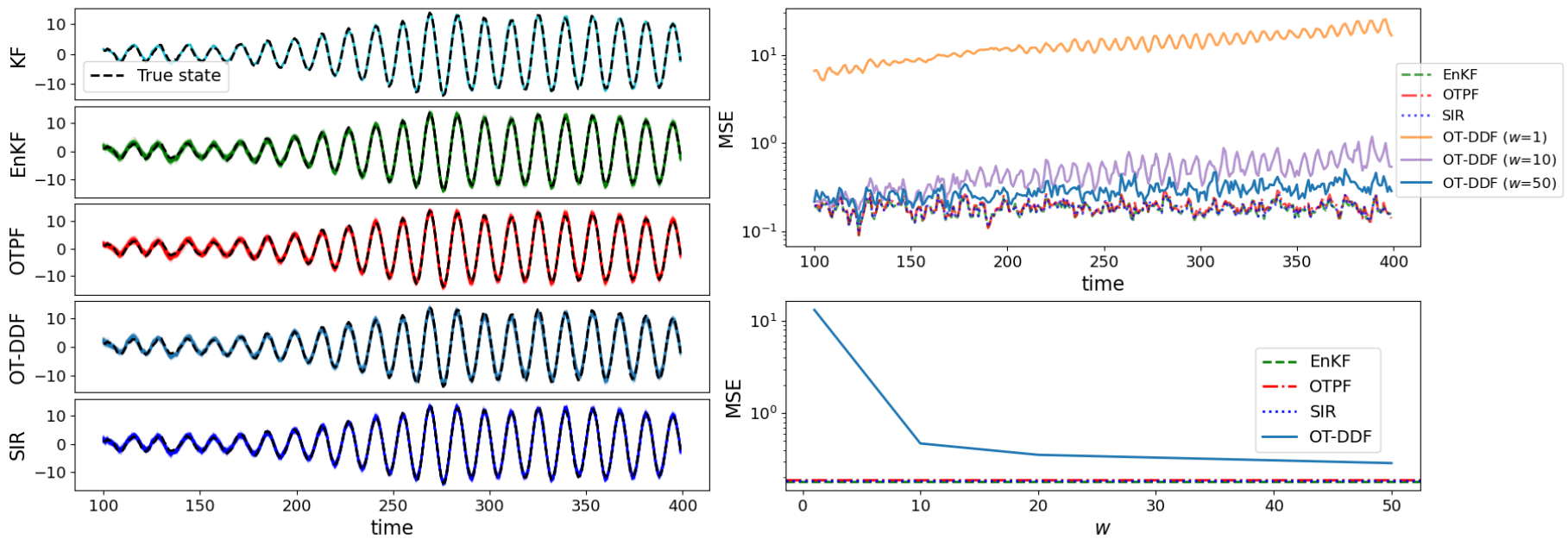
$$X_t = aX_{t-1} + \sigma V_t$$
$$Y_t = h(X_t) + \sigma W_t$$

M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Numerical example

Model:

$$X_t = aX_{t-1} + \sigma V_t$$
$$Y_t = X_t + \sigma W_t$$

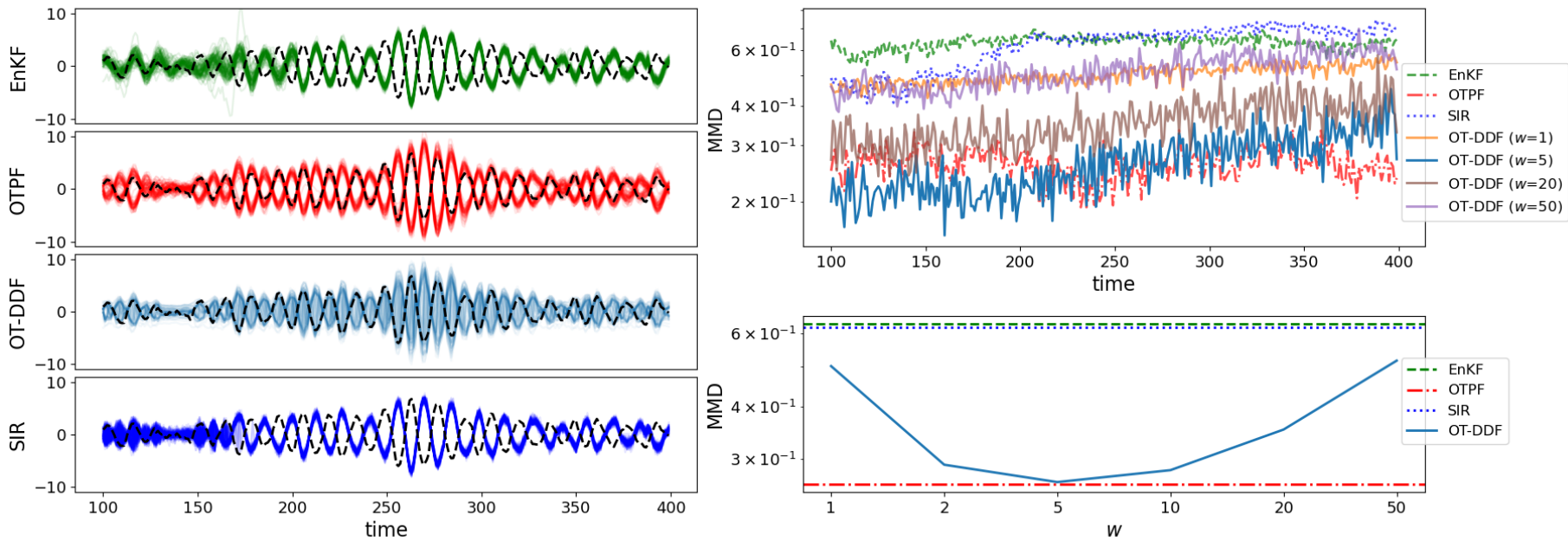


M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Numerical example

Model:

$$X_t = aX_{t-1} + \sigma V_t$$
$$Y_t = X_t^2 + \sigma W_t$$



M. Al-Jarrah, B. Hosseini, A. Taghvaei "Data-driven approximation of stationary nonlinear filters with optimal transport maps" IEEE Conference on Decision and Control (CDC). 2024.

Outline

- Foundations
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Outline

- Foundations
- Algorithmic development
  - Optimal Transport filter
  - Data-driven setting
  - Comparison to other generative models

Comparison with other generative models
Methods

Filter	Transport/coupling	Learning	Inference
Transport-based: learn a single map from prior to posterior; sampling is one-shot			
OTF	OT map T	adversarial (max min)	one-shot
KRF	KR map S	maximum likelihood	one-shot, bisection
Flow-based: learn a continuous-time dynamics; sampling integrates an ODE/SDE			
FMF	deterministic interpolant	regression	ODE, Euler steps
SIF	stochastic interpolant	regression	ODE, or SDE
SBF	fwd/bwd drifts	minimum energy	(reversed) SDE

- Optimal Transport Filter (OTF)
- Knothe–Rosenblatt rearrangement Filter (KRF)
- Flow Matching Filter (FMF)
- Stochastic Interpolant Filter (SIF)
- Schrödinger Bridge Filter (SBF)

M. Al-Jarrah, W. Deng, B. Hosseini, A. Taghvaei, "Generative models for simulation based filtering: Formulations and Empirical Comparisons", 2026.

Comparison with other generative models
Methods

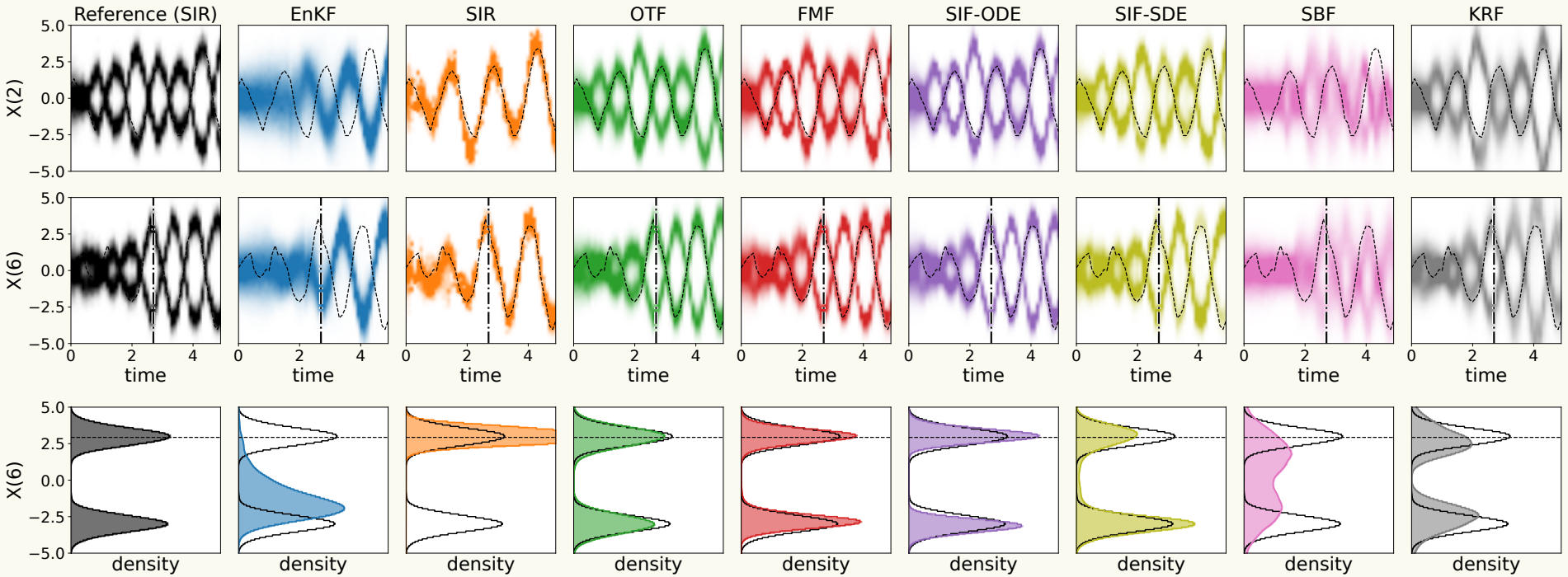
Filter	Transport/coupling	Learning	Inference
Transport-based: learn a single map from prior to posterior; sampling is one-shot			
OTF	OT map T	adversarial (max min)	one-shot
KRF	KR map S	maximum likelihood	one-shot, bisection
Flow-based: learn a continuous-time dynamics; sampling integrates an ODE/SDE			
FMF	deterministic interpolant	regression	ODE, Euler steps
SIF	stochastic interpolant	regression	ODE, or SDE
SBF	fwd/bwd drifts	minimum energy	(reversed) SDE

- Optimal Transport Filter (OTF)
- Knothe–Rosenblatt rearrangement Filter (KRF)
- Flow Matching Filter (FMF)
- Stochastic Interpolant Filter (SIF)
- Schrödinger Bridge Filter (SBF)

M. Al-Jarrah, W. Deng, B. Hosseini, A. Taghvaei, "Generative models for simulation based filtering: Formulations and Empirical Comparisons", 2026.

Comparison with other generative models

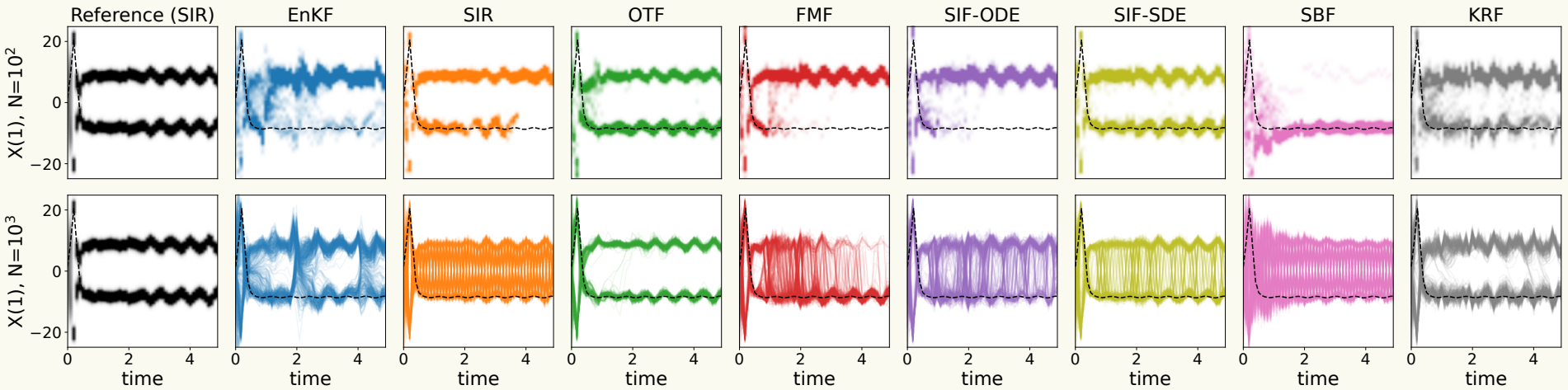
Numerical comparsion



- Representing multi-modal uncertainty
- Regularity of particle trajectories
- Scaling with dimension and accuracy-compute trade-off

M. Al-Jarrah, W. Deng, B. Hosseini, A. Taghvaei, "Generative models for simulation based filtering: Formulations and Empirical Comparisons", 2026.

Comparison with other generative models
Numerical comparsion

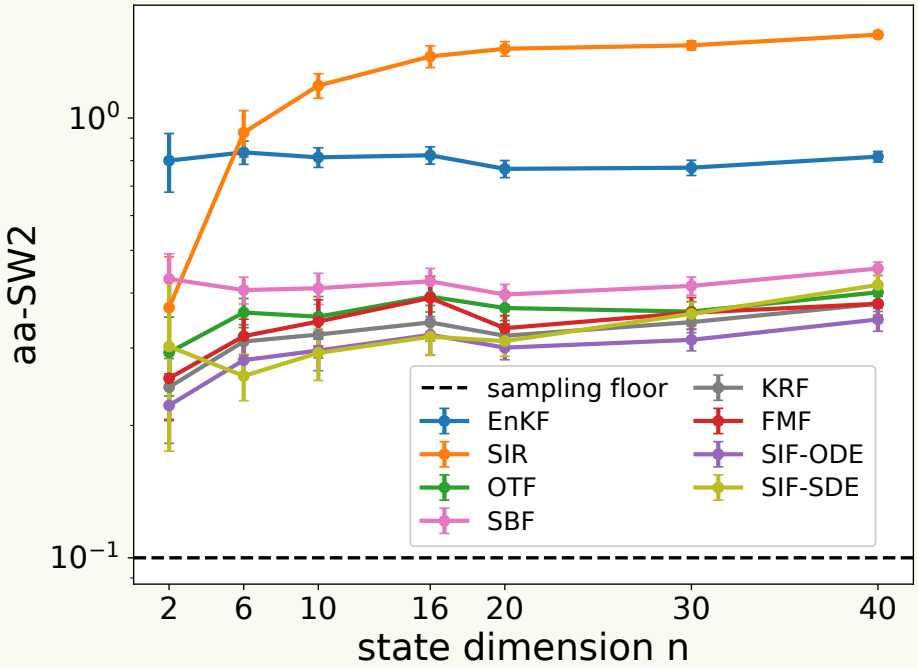


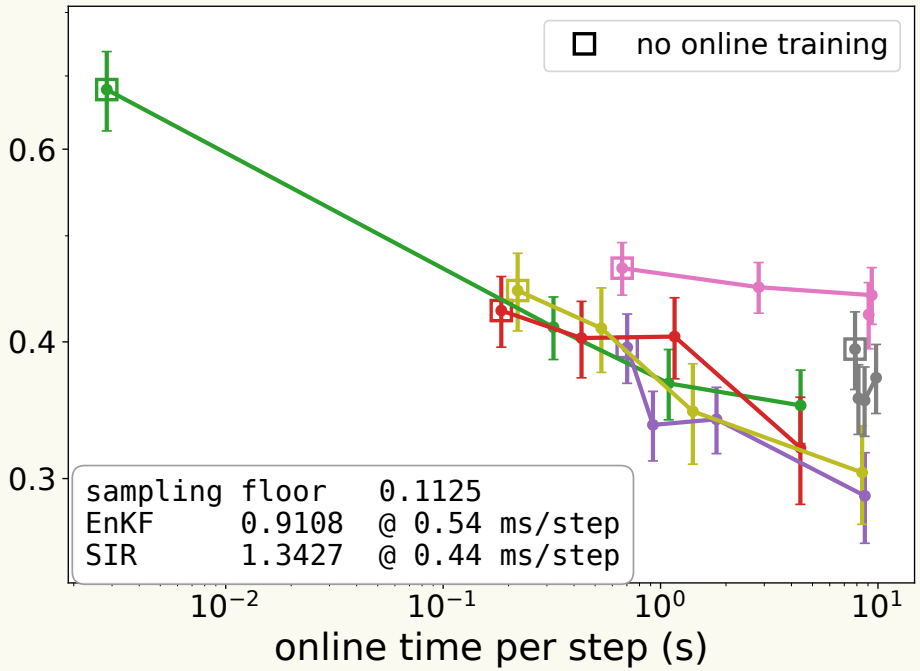
- Representing multi-modal uncertainty
- Regularity of particle trajectories
- Scaling with dimension and accuracy-compute trade-off

M. Al-Jarrah, W. Deng, B. Hosseini, A. Taghvaei, "Generative models for simulation based filtering: Formulations and Empirical Comparisons", 2026.

Comparison with other generative models

Numerical comparsion





- Representing multi-modal uncertainty
- Regularity of particle trajectories
- Scaling with dimension and accuracy-compute trade-off

M. Al-Jarrah, W. Deng, B. Hosseini, A. Taghvaei, "Generative models for simulation based filtering: Formulations and Empirical Comparisons", 2026.

Acknowledgments



M. Al-Jarrah



M. Martino



N. Jin



B. Hosseini

NSF

AFOSR

References: