Towards Data-Driven Nonlinear Filtering Algorithms

Presented at the 26th International Symposium on
Mathematical Theory of Networks and Systems, Cambridge, UK

Amirhossein Taghvaei

Department of Aeronautics & Astronautics
University of Washington, Seattle

Aug 22, 2024

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
- *Optimal Transport Particle Filters*
  Mohammad Al-Jarrah, Amirhossein Taghvaei, Bamdad Hosseini
  IEEE Conference on Decision and Control (CDC), Singapore, 2023
- Computational optimal transport and filtering on Riemannian manifolds
  D. Grange, M. Al-Jarrah, R. Baptista, A. Taghvaei, T. Georgiou, S. Phillips, A. Tannenbaum
  IEEE Control Systems Letters, 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
- *Optimal Transport Particle Filters*
  Mohammad Al-Jarrah, Amirhossein Taghvaei, Bamdad Hosseini
  IEEE Conference on Decision and Control (CDC), Singapore, 2023
- Computational optimal transport and filtering on Riemannian manifolds
  D. Grange, M. Al-Jarrah, R. Baptista, A. Taghvaei, T. Georgiou, S. Phillips, A. Tannenbaum
  IEEE Control Systems Letters, 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:** Bayes' law and its fundamental challenges
- **Part II:** Conditioning with optimal transport maps
- **Part III:** Application to nonlinear filtering
- **Part IV:** Extension to data-driven setting

Outline

- **Part I:** Bayes' law and its fundamental challenges
- **Part II:** Conditioning with optimal transport maps
- **Part III:** Application to nonlinear filtering
- **Part IV:** Extension to data-driven setting

Bayes' law

Problem:

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

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

- Data-driven setting: $P_{X,Y}$ is not available.

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

Approximate: $P_{X|Y=y}$ for any given observation y

Bayes' law

Problem:

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

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

- Data-driven setting: $P_{X,Y}$ is not available.

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

Approximate: $P_{X|Y=y}$ for any given observation y

Bayes' law

Problem:

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

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

- Data-driven setting: $P_{X,Y}$ is not available.

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

Approximate: $P_{X|Y=y}$ for any given observation y

Existing methodologies

Kalman filter (KF):

- Assumes (X, Y) is jointly Gaussian

$$P_{X,Y} = N(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} \\ \Sigma_{Y,X} & \Sigma_Y \end{bmatrix})$$

- Implements the conditioning formula for jointly Gaussian random variables

$$P_{X|Y=y} = N(m_X + K(y - m_Y), \Sigma_X - \Sigma_{X,Y}\Sigma_Y^{-1}\Sigma_{Y,X})$$

- Data-driven counterpart: Fit a Gaussian distribution to the data $(X^i, Y^i)_{i=1}^N$ and implement the conditioning formula → Ensemble Kalman filter (EnKF)
- Widely used in meteorology
- Fundamentally limited to Gaussian settings

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)

Existing methodologies

Kalman filter (KF):

- Assumes (X,Y) is jointly Gaussian

$$P_{X,Y} = N\left(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} \\ \Sigma_{Y,X} & \Sigma_Y \end{bmatrix}\right)$$

- Implements the conditioning formula for jointly Gaussian random variables

$$P_{X|Y=y} = N(m_X + K(y - m_Y), \Sigma_X - \Sigma_{X,Y}\Sigma_Y^{-1}\Sigma_{Y,X})$$

- Data-driven counterpart: Fit a Gaussian distribution to the data $(X^i, Y^i)_{i=1}^N$ and implement the conditioning formula $\rightarrow$ Ensemble Kalman filter (EnKF)
- Widely used in meteorology
- Fundamentally limited to Gaussian settings

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)

Existing methodologies

Kalman filter (KF):

- Assumes (X, Y) is jointly Gaussian

$$P_{X,Y} = N(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} \\ \Sigma_{Y,X} & \Sigma_Y \end{bmatrix})$$

- Implements the conditioning formula for jointly Gaussian random variables

$$P_{X|Y=y} = N(m_X + K(y - m_Y), \Sigma_X - \Sigma_{X,Y}\Sigma_Y^{-1}\Sigma_{Y,X})$$

- Data-driven counterpart: Fit a Gaussian distribution to the data $(X^i, Y^i)_{i=1}^N$ and implement the conditioning formula $\rightarrow$ Ensemble Kalman filter (EnKF)
- Widely used in meteorology
- Fundamentally limited to Gaussian settings

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)

Existing methodologies

Kalman filter (KF):

- Assumes (X, Y) is jointly Gaussian

$$P_{X,Y} = N(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} \\ \Sigma_{Y,X} & \Sigma_Y \end{bmatrix})$$

- Implements the conditioning formula for jointly Gaussian random variables

$$P_{X|Y=y} = N(m_X + K(y - m_Y), \Sigma_X - \Sigma_{X,Y}\Sigma_Y^{-1}\Sigma_{Y,X})$$

- Data-driven counterpart: Fit a Gaussian distribution to the data $(X^i, Y^i)_{i=1}^N$ and implement the conditioning formula $\rightarrow$ Ensemble Kalman filter (EnKF)
- Widely used in meteorology
- Fundamentally limited to Gaussian settings

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)

Existing methodologies

Kalman filter (KF):

- Assumes (X, Y) is jointly Gaussian

$$P_{X,Y} = N\left(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} \\ \Sigma_{Y,X} & \Sigma_Y \end{bmatrix}\right)$$

- Implements the conditioning formula for jointly Gaussian random variables

$$P_{X|Y=y} = N(m_X + K(y - m_Y), \Sigma_X - \Sigma_{X,Y}\Sigma_Y^{-1}\Sigma_{Y,X})$$

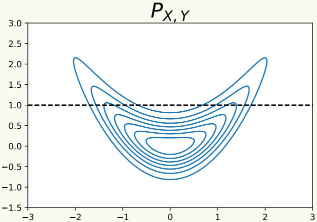
- Data-driven counterpart: Fit a Gaussian distribution to the data $(X^i, Y^i)_{i=1}^N$ and implement the conditioning formula $\rightarrow$ Ensemble Kalman filter (EnKF)
- Widely used in meteorology
- Fundamentally limited to Gaussian settings

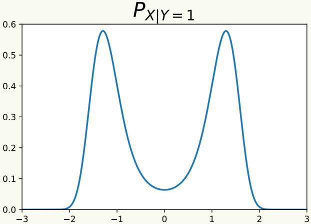
G. Evensen. "Data Assimilation. The Ensemble Kalman Filter" (2006)
S. Reich, "A dynamical systems framework for intermittent data assimilation" (2011)
E. Calvello, S. Reich, and A. M. Stuart, "Ensemble kalman methods: a mean field perspective" (2022)

Setup:

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

EnKF:





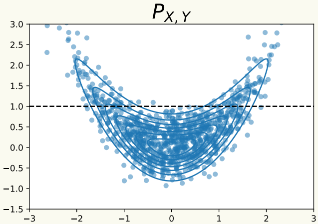
Illustrative example
Fundamental challenges of EnKF

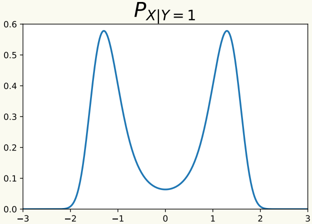
Setup:

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

EnKF:

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



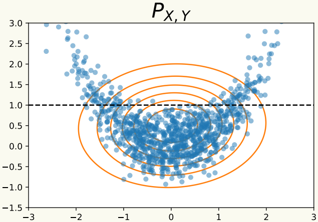


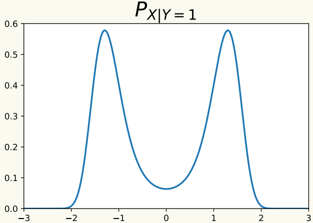
Setup:

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

EnKF:

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



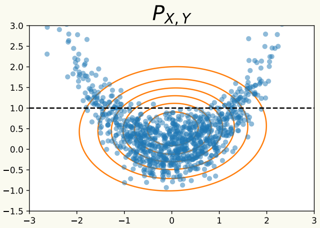


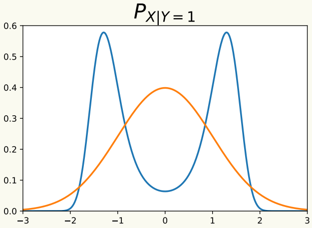
Setup:

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

EnKF:

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





Existing methodologies

Importance sampling (IS) particle filter:

- Requires samples/particles $(X^i)_{i=1}^N \overset{\text{i.i.d}}{\sim} P_X$ and likelihood function $P_{Y|X}$
- Compute the weights

$$w^i \propto P_{Y=y|X=X^i}$$

- Approximate the posterior as weighted empirical distribution

$$P_{X|Y=y} \approx \sum_{i=1}^N w^i \delta_{X^i}$$

- Asymptotically exact as $N \to \infty$
- Suffers from weight degeneracy issue

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)

Importance sampling (IS) particle filter:

- Requires samples/particles $(X^i)_{i=1}^N \overset{\text{i.i.d}}{\sim} P_X$ and likelihood function $P_{Y|X}$
- Compute the weights

$$w^i \propto P_{Y=y|X=X^i}$$

- Approximate the posterior as weighted empirical distribution

$$P_{X|Y=y} \approx \sum_{i=1}^N w^i \delta_{X^i}$$

- Asymptotically exact as $N \to \infty$
- Suffers from weight degeneracy issue

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)

Existing methodologies

Importance sampling (IS) particle filter:

- Requires samples/particles $(X^i)_{i=1}^N \overset{\text{i.i.d}}{\sim} P_X$ and likelihood function $P_{Y|X}$
- Compute the weights

$$w^i \propto P_{Y=y|X=X^i}$$

- Approximate the posterior as weighted empirical distribution

$$P_{X|Y=y} \approx \sum_{i=1}^{N} w^i \delta_{X^i}$$

- Asymptotically exact as $N \to \infty$
- Suffers from weight degeneracy issue

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)

Existing methodologies

Importance sampling (IS) particle filter:

- Requires samples/particles $(X^i)_{i=1}^N \overset{\text{i.i.d}}{\sim} P_X$ and likelihood function $P_{Y|X}$
- Compute the weights

$$w^i \propto P_{Y=y|X=X^i}$$

- Approximate the posterior as weighted empirical distribution

$$P_{X|Y=y} \approx \sum_{i=1}^N w^i \delta_{X^i}$$

- Asymptotically exact as $N \to \infty$
- Suffers from weight degeneracy issue

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)

Importance sampling (IS) particle filter:

- Requires samples/particles $(X^i)_{i=1}^N \overset{\text{i.i.d}}{\sim} P_X$ and likelihood function $P_{Y|X}$
- Compute the weights

$$w^i \propto P_{Y=y|X=X^i}$$

- Approximate the posterior as weighted empirical distribution

$$P_{X|Y=y} \approx \sum_{i=1}^N w^i \delta_{X^i}$$

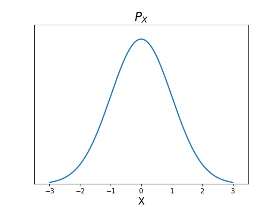
- Asymptotically exact as $N \to \infty$
- Suffers from weight degeneracy issue

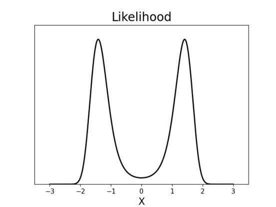
P. Del Moral, A.Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. (2001)
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).
P. Rebeschini and R. Van Handel, Can local particle filters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

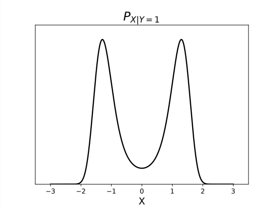
Example:

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

Importance sampling (IS):







small noise regime: $\epsilon \to 0$

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

Illustrative example

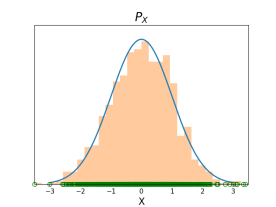
Fundamental challenges of importance sampling

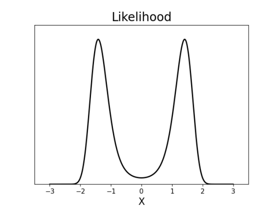
Example:

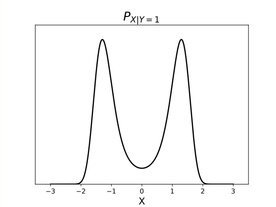
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

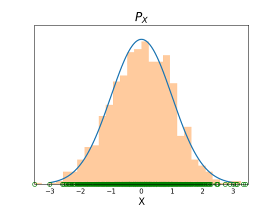
This is the main reason for the curse of dimensionality of IS-based particle filters

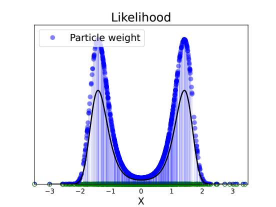
Example:

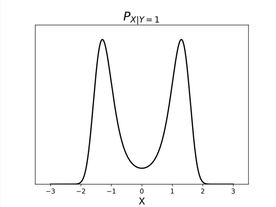
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

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

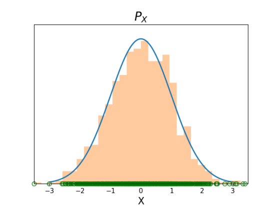
Illustrative example
Fundamental challenges of importance sampling

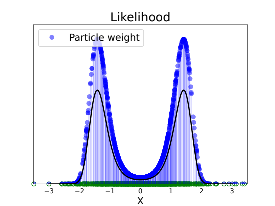
Example:

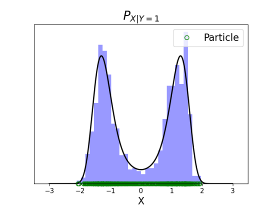
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

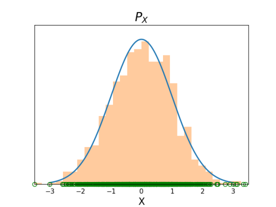
This is the main reason for the curse of dimensionality of IS-based particle filters

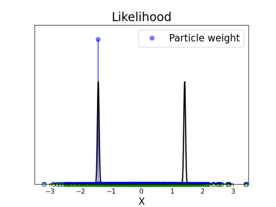
Example:

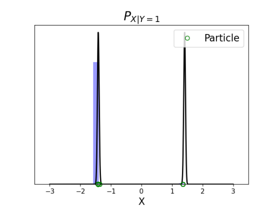
- $X \sim \mathcal{N}(0,1)$
- $Y = \frac{1}{2}X^2 + \epsilon W$
- $P_{X|Y=1} = ?$

Importance sampling (IS):

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







small noise regime: $\epsilon \to 0$

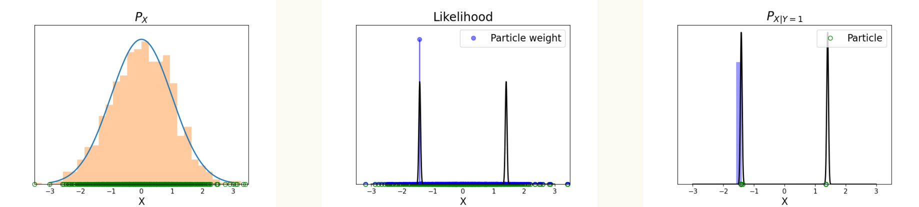
This is the main reason for the curse of dimensionality of IS-based particle filters

Example:

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

Importance sampling (IS):

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



small noise regime: $\epsilon \to 0$

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

Curse of dimensionality in particle filters

- $X, Y \in \mathbb{R}^n$ with i.i.d. components.
- Exact posterior: π_{exact}
- IS approximation: $\pi_{\text{IS}}^{(N)}$
- Asymptotic limit as $N \to \infty$:

$$d(\pi_{\text{exact}}, \pi_{\text{IS}}^{(N)}) \simeq C \frac{\gamma^n}{\sqrt{N}}$$

where $d(\cdot, \cdot)$ is the dual bounded metric and $\gamma > 1$.

- Good news: accurate as $N \to \infty$ (universal for any prior and likelihood)
- Bad news: error scales exponentially with the dimension n
- Remedy: exploit problem specific properties (e.g. spatial correlation decay in localization methods)
- Alternative method: replacing IS with control or coupling-based techniques

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)

Curse of dimensionality in particle filters

- $X, Y \in \mathbb{R}^n$ with i.i.d. components.
- Exact posterior: π_{exact}
- IS approximation: $\pi_{\text{IS}}^{(N)}$
- Asymptotic limit as $N \to \infty$:

$$d(\pi_{\text{exact}}, \pi_{\text{IS}}^{(N)}) \simeq C \frac{\gamma^n}{\sqrt{N}}$$

where $d(\cdot, \cdot)$ is the dual bounded metric and $\gamma > 1$.

- Good news: accurate as $N \to \infty$ (universal for any prior and likelihood)
- Bad news: error scales exponentially with the dimension n
- Remedy: exploit problem specific properties (e.g. spatial correlation decay in localization methods)
- Alternative method: replacing IS with control or coupling-based techniques

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)

Curse of dimensionality in particle filters

- $X, Y \in \mathbb{R}^n$ with i.i.d. components.
- Exact posterior: π_{exact}
- IS approximation: $\pi_{\text{IS}}^{(N)}$
- Asymptotic limit as $N \to \infty$:

$$d(\pi_{\text{exact}}, \pi_{\text{IS}}^{(N)}) \simeq C \frac{\gamma^n}{\sqrt{N}}$$

  where $d(\cdot, \cdot)$ is the dual bounded metric and $\gamma > 1$.

- Good news: accurate as $N \to \infty$ (universal for any prior and likelihood)
- Bad news: error scales exponentially with the dimension n
- Remedy: exploit problem specific properties (e.g. spatial correlation decay in localization methods)
- Alternative method: replacing IS with control or coupling-based techniques

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)

Curse of dimensionality in particle filters

- $X, Y \in \mathbb{R}^n$ with i.i.d. components.
- Exact posterior: π_{exact}
- IS approximation: $\pi_{\text{IS}}^{(N)}$
- Asymptotic limit as $N \to \infty$:

$$d(\pi_{\text{exact}}, \pi_{\text{IS}}^{(N)}) \simeq C \frac{\gamma^n}{\sqrt{N}}$$

where $d(\cdot,\cdot)$ is the dual bounded metric and $\gamma > 1$.

- Good news: accurate as $N \to \infty$ (universal for any prior and likelihood)
- Bad news: error scales exponentially with the dimension n
- Remedy: exploit problem specific properties (e.g. spatial correlation decay in localization methods)
- Alternative method: replacing IS with control or coupling-based techniques

---

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)

Curse of dimensionality in particle filters

- $X, Y \in \mathbb{R}^n$ with i.i.d. components.
- Exact posterior: π_{exact}
- IS approximation: $\pi_{\mathsf{IS}}^{(N)}$
- Asymptotic limit as $N \to \infty$:

$$d(\pi_{\mathsf{exact}}, \pi_{\mathsf{IS}}^{(N)}) \simeq C \frac{\gamma^n}{\sqrt{N}}$$

  where $d(\cdot, \cdot)$ is the dual bounded metric and $\gamma > 1$.
- Good news: accurate as $N \to \infty$ (universal for any prior and likelihood)
- Bad news: error scales exponentially with the dimension n
- Remedy: exploit problem specific properties (e.g. spatial correlation decay in localization methods)
- Alternative method: replacing IS with control or coupling-based techniques

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)

Curse of dimensionality in particle filters

- $X, Y \in \mathbb{R}^n$ with i.i.d. components.
- Exact posterior: π_{exact}
- IS approximation: $\pi_{\text{IS}}^{(N)}$
- Asymptotic limit as $N \to \infty$:

$$d(\pi_{\text{exact}}, \pi_{\text{IS}}^{(N)}) \simeq C \frac{\gamma^n}{\sqrt{N}}$$

where $d(\cdot, \cdot)$ is the dual bounded metric and $\gamma > 1$.

- Good news: accurate as $N \to \infty$ (universal for any prior and likelihood)
- Bad news: error scales exponentially with the dimension n
- Remedy: exploit problem specific properties (e.g. spatial correlation decay in localization methods)
- Alternative method: replacing IS with control or coupling-based techniques

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]
- 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]
- ...

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]
- 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]
- ...

This talk: Conditioning with optimal transport map

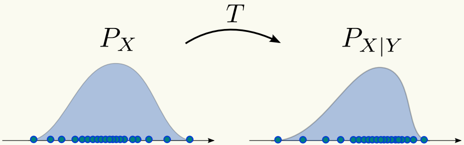
Outline

- **Part I:** Bayes' law and its fundamental challenges
- **Part II:** Conditioning with optimal transport maps
- **Part II:** Application to nonlinear filtering
- **Part III:** Extension to data-driven setting

Outline

- **Part I:** Bayes' law and its fundamental challenges
- **Part II:** Conditioning with optimal transport maps
- **Part II:** Application to nonlinear filtering
- **Part III:** Extension to data-driven setting

Conditioning with transport maps

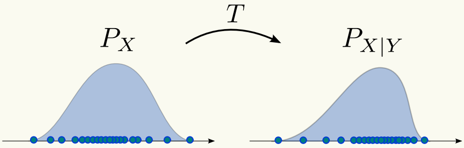


$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

Example:

Questions: In a general setting,

Conditioning with transport maps



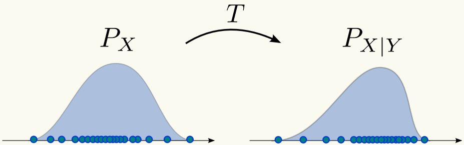
$$X^i \sim P_X \longrightarrow T(X^i, y) \sim P_{X|Y=y}$$

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,

Conditioning with transport maps



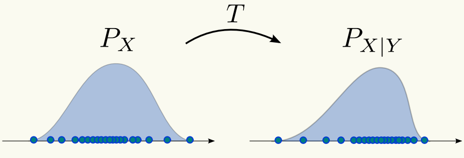
$$X^i \sim P_X \;\longrightarrow\; T(X^i, y) \sim P_{X|Y=y}$$

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,

Conditioning with transport maps



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

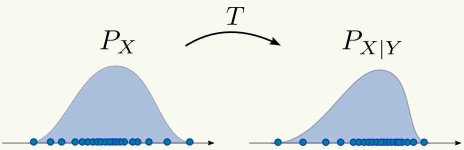
Example:

- 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



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

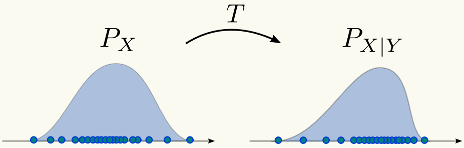
Example:

- 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



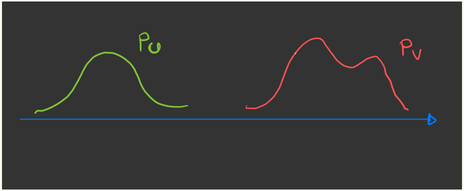
$$X^i \sim P_X \ \longrightarrow \ T(X^i, y) \sim P_{X|Y=y}$$

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 $X \mapsto X + K(y - Y)$

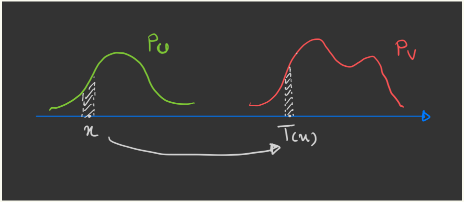
Questions: In a general setting,

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



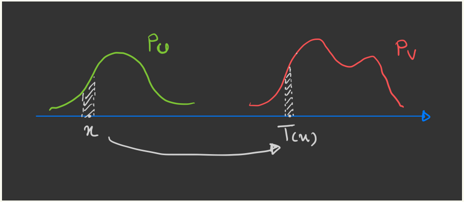
- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_{\#}P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:



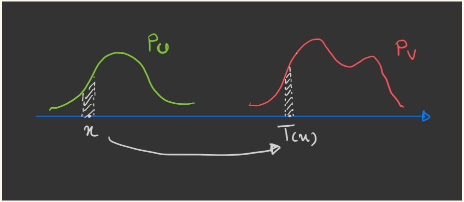
- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_\# P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:



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

Questions:



- 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 $\|T(x) - x\|^2$

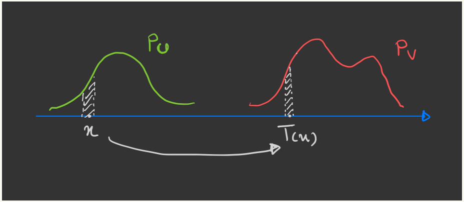
Questions:

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

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

Background on optimal transportation theory

Monge problem and Brenier's result

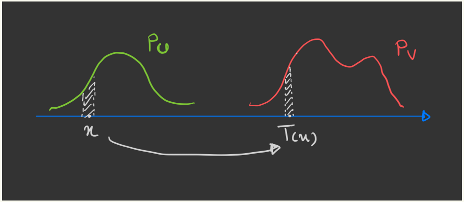


- Given two random variables $U \sim P_U$ and $V \sim P_V$
- find a map $x \mapsto T(x)$ that transports P_U to P_V, i.e. $T_{\#} P_U = P_V$
- with minimal transportation cost $\|T(x) - x\|^2$

Questions:

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

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



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

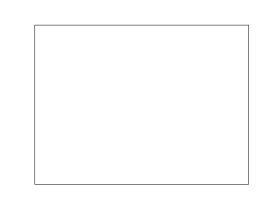
Questions:

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

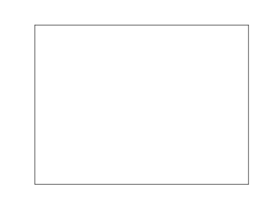
$$\max_{f \in c\text{-concave}} \min_{T} \mathbb{E}\left[\frac{1}{2}\|T(U) - U\|^2 - f(T(U)) + f(V)\right]$$

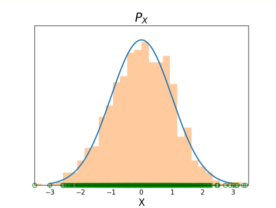
Conditioning with optimal transport map

Illustrative example

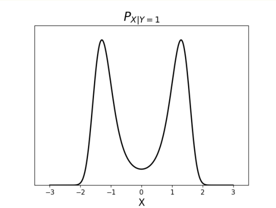


$\rightarrow$

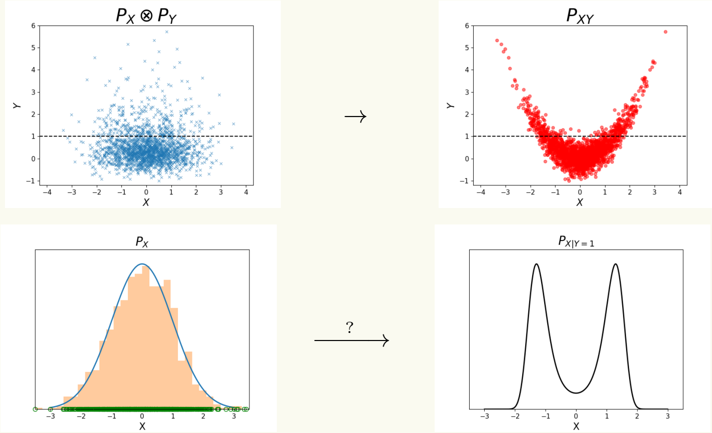




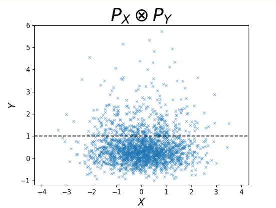
$\xrightarrow{?}$



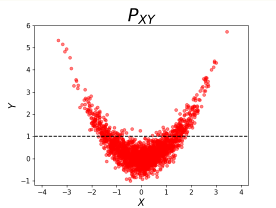
Conditioning with optimal transport map
Illustrative example

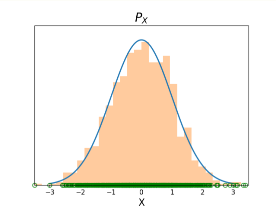


Conditioning with optimal transport map
Illustrative example

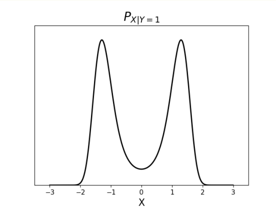


$\xrightarrow{(T(X,Y),Y)}$



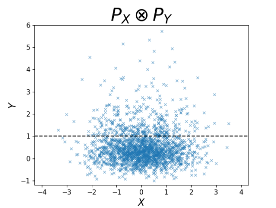


$\xrightarrow{?}$

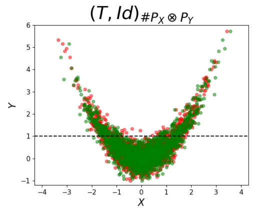


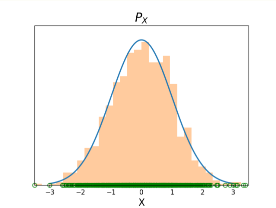
Conditioning with optimal transport map

Illustrative example

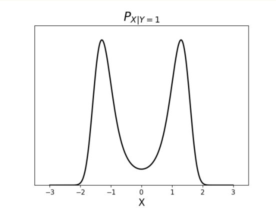


$\xrightarrow{(T(X,Y),Y)}$



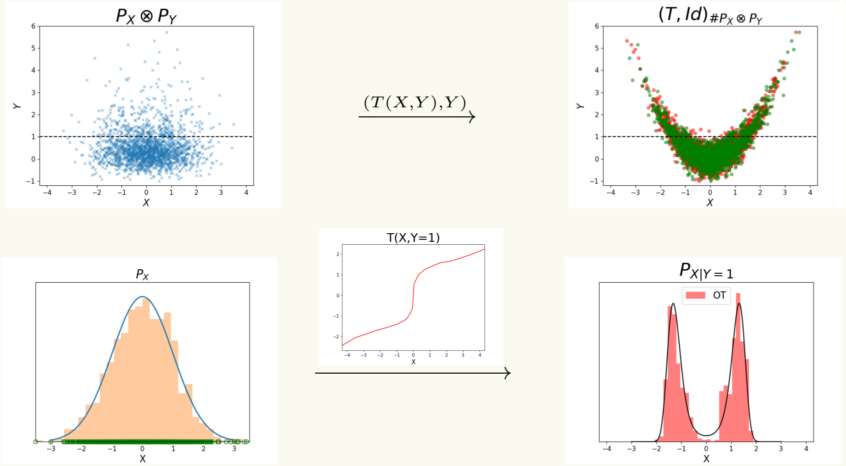


$\xrightarrow{?}$



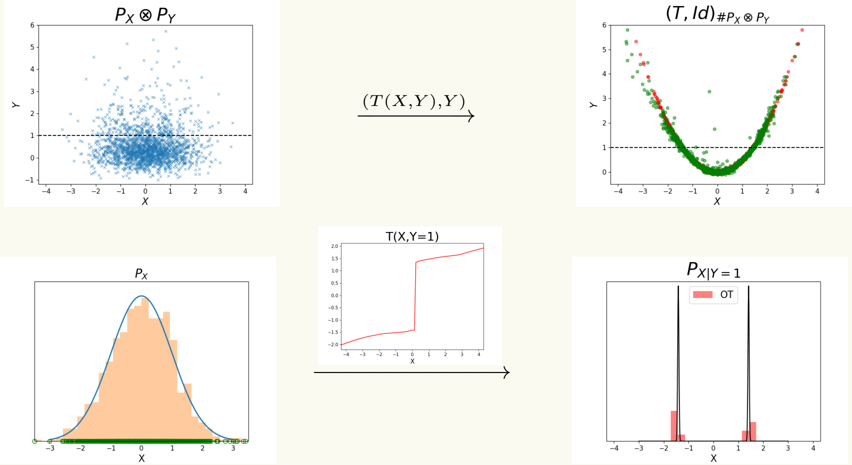
Conditioning with optimal transport map

Illustrative example



Conditioning with optimal transport map

Illustrative example



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 max-min formulation:

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

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)

Conditioning with optimal transport map

Variational formulation of the Bayes' law

Bayes law:

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

Conditional max-min formulation:

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

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)

Conditioning with optimal transport map

Variational formulation of the Bayes' law

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

Conditional max-min formulation:

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

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)

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

Conditional max-min formulation:

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

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)

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

(Conditional) Brenier's theorem

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

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

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

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

B. Hosseini, A. Hsu, A. Taghvaei Conditional Optimal Transport on Function Spaces (2023)
G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal trans- port approach.v (2016).

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

(Conditional) Brenier's theorem

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

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

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

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

B. Hosseini, A. Hsu, A. Taghvaei Conditional Optimal Transport on Function Spaces (2023)
G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal trans- port approach.v (2016).

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

(Conditional) Brenier's theorem

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

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

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

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

B. Hosseini, A. Hsu, A. Taghvaei Conditional Optimal Transport on Function Spaces (2023)
G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal trans- port approach.v (2016).

About convexity assumption

- Ensuring the assumption that

$$x \mapsto \frac{1}{2}\|x\|^2 - f(x,y) \text{ is } \alpha\text{-strongly convex for all } y$$

is computationally challenging
- In practice, we do not enforce a convexity constraint
- The optimizer outputs f that is, sometimes, slightly non-convex

About convexity assumption

- Ensuring the assumption that

$$x \mapsto \frac{1}{2}\|x\|^2 - f(x,y) \text{ is } \alpha\text{-strongly convex for all } y$$

  is computationally challenging

- In practice, we do not enforce a convexity constraint
- The optimizer outputs f that is, sometimes, slightly non-convex

About convexity assumption

- Ensuring the assumption that

$$x \mapsto \frac{1}{2}\|x\|^2 - f(x,y) \text{ is } \alpha\text{-strongly convex for all } y$$

  is <u>computationally challenging</u>
- In practice, we do not enforce a convexity constraint
- The optimizer outputs f that is, sometimes, slightly non-convex

About convexity assumption

- Ensuring the assumption that

$$x \mapsto \frac{1}{2}\|x\|^2 - f(x,y) \text{ is } \alpha\text{-strongly convex for all } y$$

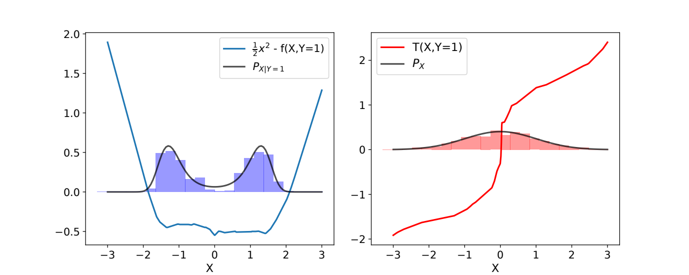
  is computationally challenging
- In practice, we do not enforce a convexity constraint
- The optimizer outputs f that is, sometimes, slightly non-convex

About convexity assumption

- Ensuring the assumption that

$$x \mapsto \frac{1}{2}\|x\|^2 - f(x,y) \text{ is } \alpha\text{-strongly convex for all } y$$

  is <u>computationally challenging</u>
- In practice, we do not enforce a convexity constraint
- The optimizer outputs f that is, sometimes, slightly non-convex



Outline

- **Part I:** Bayes' law and its fundamental challenges
- **Part II:** Conditioning with optimal transport maps
- **Part II:** Application to nonlinear filtering
- **Part III:** Extension to data-driven setting

Outline

- **Part I:** Bayes' law and its fundamental challenges
- **Part II:** Conditioning with optimal transport maps
- **Part II:** Application to nonlinear filtering
- **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

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

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

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

Nonlinear filtering: numerical approximation of the posterior π_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

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: numerical approximation of the posterior π_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

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

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

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

Nonlinear filtering: numerical approximation of the posterior π_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

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

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

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

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

Filtering equations

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

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

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

- Recursive update law for the posterior

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

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

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

Filtering equations

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

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

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

- Recursive update law for the posterior

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

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

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

Filtering equations

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

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

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

- Recursive update law for the posterior

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

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

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

Filtering equations

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

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

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

- Recursive update law for the posterior

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

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

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

Optimal Transport Filter

Filter design steps:

exact posterior: $\pi_{t-1} \longrightarrow \pi_{t|t-1} = \mathcal{A}\pi_{t-1} \longrightarrow \pi_t = \mathcal{B}_{Y_t}(\pi_{t|t-1})$

mean-field process: $\bar{X}_{t-1} \longrightarrow \bar{X}_{t|t-1} \sim a(\cdot \mid \bar{X}_{t-1}) \longrightarrow \bar{X}_t = \overline{T}_t(\bar{X}_{t|t-1}, Y_t)$

particle system: $X^i_{t-1} \longrightarrow X^i_{t|t-1} \sim a(\cdot \mid X^i_{t-1}) \longrightarrow X^i_t = \hat{T}_t(X^i_{t|t-1}, Y_t)$

Variational problem:

$$\leftarrow \max_{f} \min_{T} J(f,T; P_{\bar{X}_{t|t-1}, Y_{t|t-1}})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J\Big(f,T; \frac{1}{N}\sum_{i=1}^N \delta_{(X^i_{t|t-1}, Y^i_{t|t-1})}\Big)$$

Posterior approximation:

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

Optimal Transport Filter

Filter design steps:

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

mean-field process: $\bar{X}_{t-1} \longrightarrow \bar{X}_{t|t-1} \sim a(\cdot \mid \bar{X}_{t-1}) \longrightarrow \bar{X}_t = \overline{T}_t(\bar{X}_{t|t-1}, Y_t)$

particle system: $X^i_{t-1} \longrightarrow X^i_{t|t-1} \sim a(\cdot \mid X^i_{t-1}) \longrightarrow X^i_t = \hat{T}_t(X^i_{t|t-1}, Y_t)$

Variational problem:

$$\overline{T}_t \leftarrow \max_f \min_T \; J(f, T; P_{X_t, Y_t | Y_{1:t-1}})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} \; J(f, T; \frac{1}{N}\sum_{i=1}^{N} \delta_{(X^i_{t|t-1}, Y^i_{t|t-1})})$$

Posterior approximation:

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

Optimal Transport Filter

Filter design steps:

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

mean-field process: $\bar{X}_{t-1} \longrightarrow \bar{X}_{t|t-1} \sim a(\cdot \mid \bar{X}_{t-1}) \longrightarrow \bar{X}_t = \overline{T}_t(\bar{X}_{t|t-1}, Y_t)$

particle system: $X^i_{t-1} \longrightarrow X^i_{t|t-1} \sim a(\cdot \mid X^i_{t-1}) \longrightarrow X^i_t = \hat{T}_t(X^i_{t|t-1}, Y_t)$

Variational problem:

$$\overline{T}_t \leftarrow \max_f \min_T J(f, T; P_{X_t, Y_t | Y_{1:t-1}})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} J(f, T; \frac{1}{N}\sum_{i=1}^N \delta_{(X^i_{t|t-1}, Y^i_{t|t-1})})$$

Posterior approximation:

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

Optimal Transport Filter

Filter design steps:

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

mean-field process: $\bar{X}_{t-1} \longrightarrow \bar{X}_{t|t-1} \sim a(\cdot \mid \bar{X}_{t-1}) \longrightarrow \bar{X}_t = \overline{T}_t(\bar{X}_{t|t-1}, Y_t)$

particle system: $X^i_{t-1} \longrightarrow X^i_{t|t-1} \sim a(\cdot \mid X^i_{t-1}) \longrightarrow X^i_t = \hat{T}_t(X^i_{t|t-1}, Y_t)$

Variational problem:

$$\overline{T}_t \leftarrow \max_f \min_T \; J(f, T; P_{X_t, Y_t | Y_{1:t-1}})$$

$$\hat{T}_t \leftarrow \max_{f \in \mathcal{F}} \min_{T \in \mathcal{T}} \; J(f, T; \frac{1}{N} \sum_{i=1}^N \delta_{(X^i_{t|t-1}, Y^i_{t|t-1})})$$

Posterior approximation:

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

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:

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:

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:

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:

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

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

Theorem

Assume

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

Then,

$$d(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t) \leq C\left(\sqrt{\frac{2}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap $\epsilon_{\mathcal{F},\mathcal{T},N}$ has the decomposition

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

Theorem

Assume

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

Then,

$$d\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t\right) \leq C\left(\sqrt{\frac{2}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap $\epsilon_{\mathcal{F},\mathcal{T},N}$ has the decomposition

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

Theorem

Assume

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

Then,

$$d\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i},\pi_t\Big) \le C\left(\sqrt{\frac{2}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap $\epsilon_{\mathcal{F},\mathcal{T},N}$ has the decomposition

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

Theorem

Assume

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

Then,

$$d\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t\Big) \leq C\left(\sqrt{\frac{2}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap $\epsilon_{\mathcal{F},\mathcal{T},N}$ has the decomposition

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

Theorem

Assume

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

Then,

$$d\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i},\pi_t\Big) \leq C\left(\sqrt{\frac{2}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap $\epsilon_{\mathcal{F},\mathcal{T},N}$ has the decomposition

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

Theorem

Assume

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

Then,

$$d\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}, \pi_t\right) \leq C\left(\sqrt{\frac{2}{\alpha}\epsilon_{\mathcal{F},\mathcal{T},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t.$$

- Optimality gap $\epsilon_{\mathcal{F},\mathcal{T},N}$ has the decomposition

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

Optimal Transport Filter
Numerical example

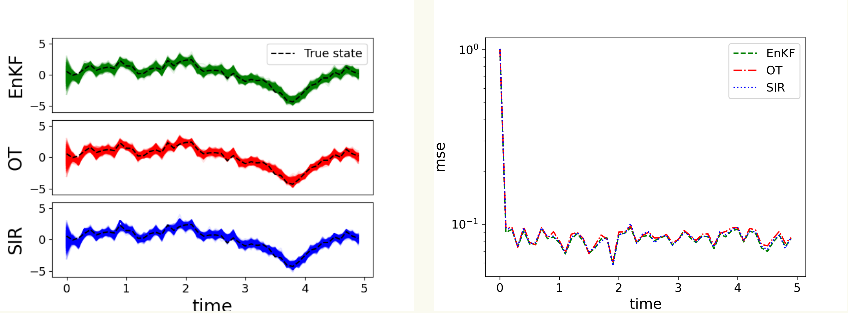
$$X_t = (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),$$
$$Y_t = X_t + \sigma_W W_t,$$

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

Optimal Transport Filter

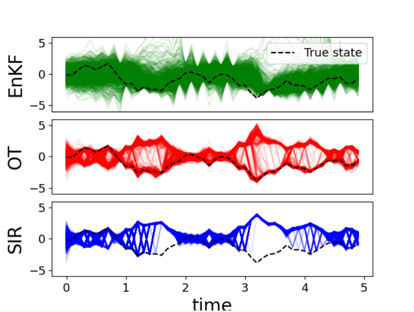
Numerical example

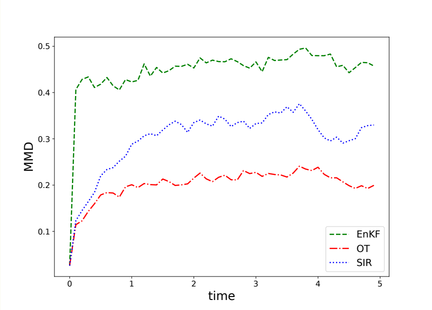
$$X_t = (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),$$
$$Y_t = X_t + \sigma_W W_t,$$



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

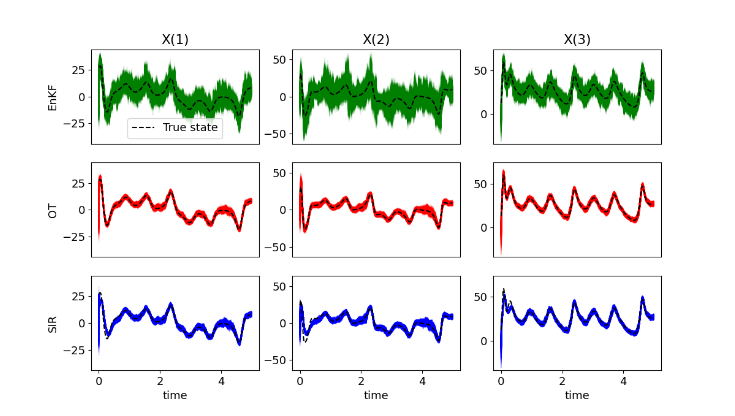
$$X_t = (1-\alpha)X_{t-1} + \sigma_V V_t, \quad X_0 \sim \mathcal{N}(0, I_n),$$
$$Y_t = X_t^2 + \sigma_W W_t,$$



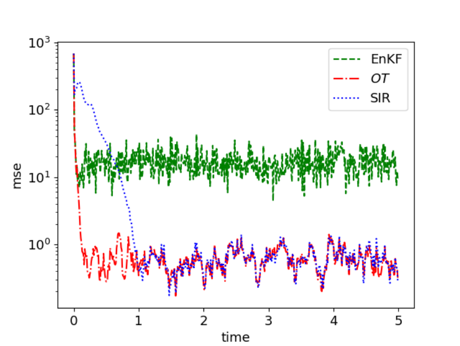


- Ensemble Kalman filter (EnKF)
- sequential importance re-sampling (SIR)
- Optimal Transport (OT)

Optimal Transport Filter
Numerical example: Lorenz 63



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



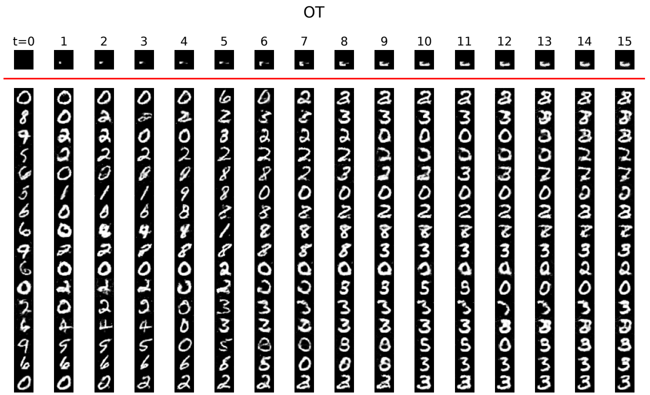
- Trajectory of the particles
- mean-squared error (mse) in estimating the state

Numerical example: Image in-painting

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

Numerical example: Image in-painting

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



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 $U(x,y) \in T_xM$

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

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

- 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

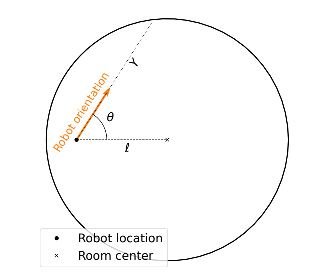
- Assume $X \in \mathcal{M}$ with metric g and geodesic distance d_g
- Replace the Euclidean distance with the geodesic distance
- Replace $T(x,y)$ with $\exp_x(U(x,y))$ where $U(x,y) \in T_x M$

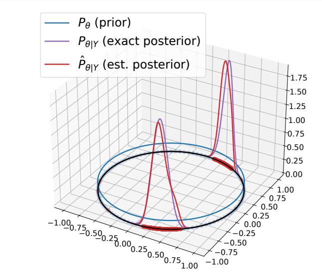
$$\max_{f:\mathcal{M}\to\mathbb{R}} \min_{U:\mathcal{M}\to T\mathcal{M}} \mathbb{E}\left[\frac{1}{2} d_g(\exp_{\bar{X}}(U(\bar{X},Y)),\bar{X})^2 - f(\exp_{\bar{X}}(U(\bar{X},Y)),Y) + f(X;Y)\right]$$

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 example: $\mathcal{M} = S^1$

- $\theta \in M$ is robot's orientation and Y is noisy measurement of distance to the wall

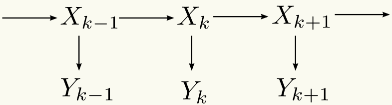




D. Grange, M. Al-Jarrah, R. Baptista, A. Taghvaei, T. Georgiou, S. Phillips, A. Tannenbaum, Computational optimal transport and filtering on Riemannian manifolds, IEEE Control Systems Letters, 2023

Summary

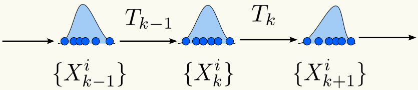
- **Mathematical model:**



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

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

- **OT approach:**



- **Variational problem:**

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

Outline

- **Part I:** Bayes' law and fundamental challenges of importance sampling
- **Part II:** Conditioning with optimal transport maps
- **Part III:** Application to nonlinear filtering
- **Part IV:** Extension to data-driven setting

Outline

- **Part I:** Bayes' law and fundamental challenges of importance sampling
- **Part II:** Conditioning with optimal transport maps
- **Part III:** Application to nonlinear filtering
- **Part IV:** Extension to data-driven setting

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 <u>unknown</u>

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:

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

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

- 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 $P_{X_w, Y_w, \dots, Y_1}$

- 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 $P_{X_w, Y_w, \ldots, Y_1}$

- 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 $P_{X_w, Y_w, \ldots, Y_1}$

- 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 $P_{X_w,Y_w,\ldots,Y_1}$

- 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 $P_{X_w, Y_w, \ldots, Y_1}$

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

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 $x \mapsto \frac{1}{2}\|x\|^2 - f(x, y_w, \ldots, y_1)$ is α-strongly convex for all $(y_w, \ldots, y_1)$.

Then,

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

Numerical example

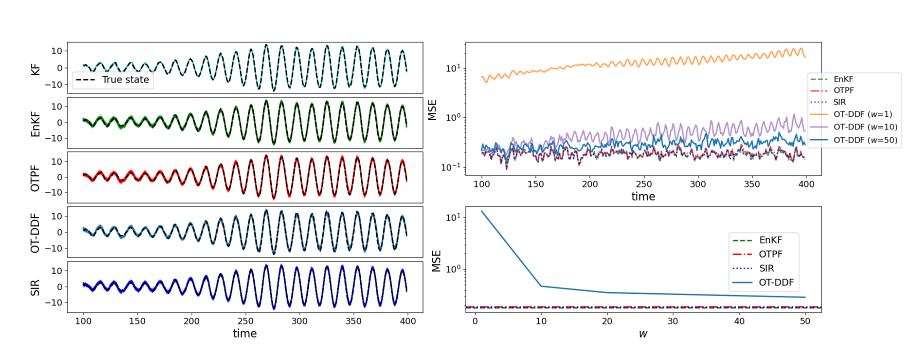
Model:

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

Numerical example

Model:

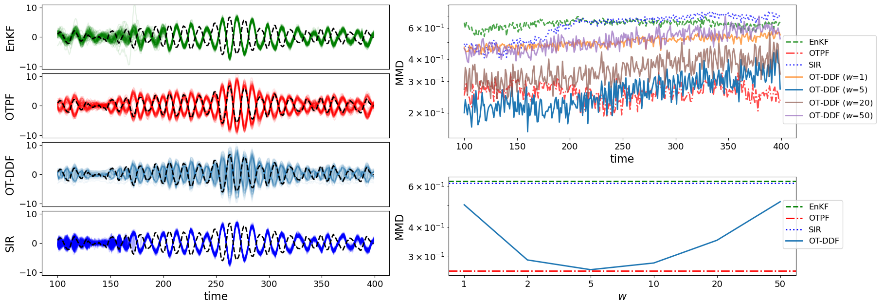
$$X_t = aX_{t-1} + \sigma V_t$$
$$Y_t = X_t + \sigma W_t$$



Numerical example

Model:

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



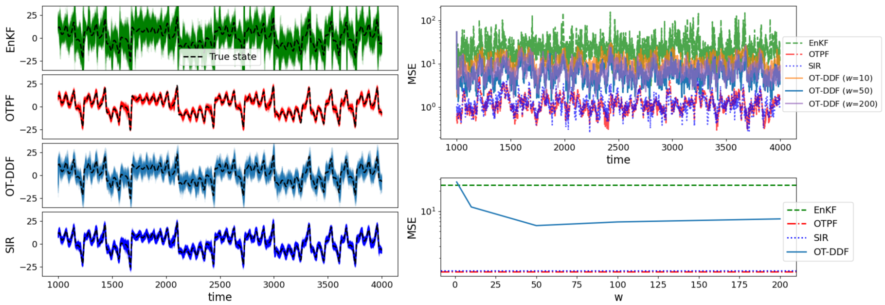
Numerical example

Lorenz 63 model

$$\dot{X} = f(X), \quad X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2 I_3),$$
$$Y_t = X_t(1) + W_t, \quad W_t \sim \mathcal{N}(0, \sigma^2), \quad \Delta t = 0.01$$

Numerical example
Lorenz 63 model

$$\dot{X} = f(X), \quad X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2 I_3),$$
$$Y_t = X_t(1) + W_t, \quad W_t \sim \mathcal{N}(0, \sigma^2), \quad \Delta t = 0.01$$



Numerical example
Lorenz 63 model

$$\dot{X} = f(X), \quad X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2 I_3),$$
$$Y_t = X_t(1) + W_t, \quad W_t \sim \mathcal{N}(0, \sigma^2), \quad \Delta t = 0.01$$

Offline training time: 46.29 seconds

One-time step update:

Method	EnKF	SIR	OTPF	OT-DDF
time	1.7×10^{-4}	2.0×10^{-4}	6.8×10^{-2}	1.5×10^{-4}

Acknowledgments

M. Al-Jarrah

N. Jin

B. Hosseini

NSF

References: