Non-linear filtering via optimal transport

Beatrice Acciaio

ETH Zurich

ongoing work with T. Schmidt (U. Freiburg)

Conference on New Monge Problems and Applications

September 14-15 2023, University Gustave Eiffel

Filtering problem

Consider the evolution of two processes in discrete time:

$$\begin{aligned} X_t &= g_t(X_{t-1}, \varepsilon_t), \quad X_0 \sim p_0 \\ Y_t &= h_t(X_t, \eta_t), \end{aligned}$$

with

- hidden (signal) process X taking value in some Polish space E
- observable process Y taking value in some Polish space F
- $(\varepsilon_t)_t$ and $(\eta_t)_t$ sequences of globally independent random variables taking value in some Polish space E' and F', respectively
- $g_t : E \times E' \to E$ and $h_t : E \times F' \to F$ measurable functions

Filtering problem

Consider the evolution of two processes in discrete time:

$$X_t = g_t(X_{t-1}, \varepsilon_t), \quad X_0 \sim p_0$$
$$Y_t = h_t(X_t, \eta_t),$$

with

- hidden (signal) process X taking value in some Polish space E
- observable process Y taking value in some Polish space F
- $(\varepsilon_t)_t$ and $(\eta_t)_t$ sequences of globally independent random variables taking value in some Polish space E' and F', respectively
- $g_t : E \times E' \to E$ and $h_t : E \times F' \to F$ measurable functions

GOAL: given the observed process (Y), infer realization of the hidden one (X):

$$\hat{X}_t = \mathbb{E}[X_t | Y_0, ..., Y_t] \quad \forall t$$

Filtering problem

A popular approach: Two-steps update.

- Propagation (prediction according to previous estimate and model dymanics):

$$\hat{X}_t^- = \mathbb{E}[X_t | Y_0, ..., Y_{t-1}] = \mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)]$$

Filtering problem

A popular approach: Two-steps update.

- Propagation (prediction according to previous estimate and model dymanics):

$$\hat{X}_t^- = \mathbb{E}[X_t | Y_0, ..., Y_{t-1}] = \mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)]$$

- Conditioning (update via Bayes rule given the observed process):

$$\hat{X}_t = \mathbb{E}[X_t | Y_0, ..., Y_{t-1}, Y_t] = \textit{function}(\hat{X}_t^-, Y_t)$$

Filtering problem

A popular approach: Two-steps update.

- Propagation (prediction according to previous estimate and model dymanics):

$$\hat{X}_t^- = \mathbb{E}[X_t | Y_0, ..., Y_{t-1}] = \mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)]$$

- Conditioning (update via Bayes rule given the observed process):

$$\hat{X}_t = \mathbb{E}[X_t | Y_0, ..., Y_{t-1}, Y_t] = \mathit{function}(\hat{X}_t^-, Y_t)$$

Main idea of our approach: use the two-step update, performing perform step 2 with a variational representation of Bayes' rule via optimal transport

Kalman filter: basic idea

- Let $(X, Y) \sim \mathcal{N}(\mu, \Sigma) \implies \xi = \frac{X-\mu_1}{\sigma_1}, \gamma = \frac{Y-\mu_2}{\sigma_2} \sim \mathcal{N}(0,1)$ with correlation $\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$
- Then

$$\xi = \rho\gamma + \sqrt{1-\rho^2}\gamma', \quad \text{with } \gamma' \sim \mathcal{N}(0,1), \text{ independent of } \gamma$$

- That is

$$X = \mu_1 + \rho \cdot \sigma_1 \frac{Y-\mu_2}{\sigma_2} + \sigma_1 \sqrt{1-\rho^2}\gamma'$$

$$X|Y \sim \mathcal{N}\left(\mu_1 + \rho \cdot \sigma_1 \frac{Y-\mu_2}{\sigma_2}, \sigma_1 \sqrt{1-\rho^2}\right)$$

- In particular,

$$\mathbb{E}[X|Y] = \mu_1 + \rho \cdot \sigma_1 \frac{Y-\mu_2}{\sigma_2}$$

Kalman filter

- Consider the system:

$$\begin{aligned} X_t &= a_t X_{t-1} + b_t \varepsilon_t, \\ Y_t &= A_t X_t + B_t \eta_t, \end{aligned}$$

with ε_t, η_t independent standard normal

Kalman filter

- Consider the system:

$$X_t = a_t X_{t-1} + b_t \varepsilon_t,$$
$$Y_t = A_t X_t + B_t \eta_t,$$

  with ε_t, η_t independent standard normal
- Then the two-steps update is given by:

$$\hat{X}_t^- = a_t \hat{X}_{t-1}$$
$$\hat{X}_t = a_t \hat{X}_{t-1} + G_t \cdot (Y_t - A_t a_t \hat{X}_{t-1})$$

  with $G_t = \frac{A_t C_t}{A_t^2 C_t + B_t^2}$ and $C_t = a_t^2 (1 - G_{t-1} A_t) C_{t-1} + b_t^2$
  (Explicit formulation of posterior distribution in a linear Gaussian setting)

Conditional expectations as transports

Lemma

Let E, F be Polish spaces, X, Y non-atomic r.v.'s taking values in E, F, resp. Then:

(i) *There exists a measurable map $T : E \times F \to E$ s.t., for $\tilde{X} \sim X, \tilde{Y} \sim Y, \tilde{X} \perp \tilde{Y}$,*

$$\left(T(\tilde{X}, \tilde{Y}), \tilde{Y}\right) \stackrel{Law}{=} (X, Y).$$

This means that $S : (x, y) \mapsto (T(x, y), y)$ is a Monge map that transports the independent coupling $P_X \otimes P_Y$ into the joint distribution P_{XY}:

$$S_{\#}(P_X \otimes P_Y) = P_{XY}.$$

Conditional expectations as transports

Lemma

Let E, F be Polish spaces, X, Y non-atomic r.v.'s taking values in E, F, resp. Then:

(i) *There exists a measurable map $T : E \times F \to E$ s.t., for $\tilde{X} \sim X, \tilde{Y} \sim Y, \tilde{X} \perp \tilde{Y}$,*

$$\big(T(\tilde{X}, \tilde{Y}), \tilde{Y}\big) \overset{Law}{=} (X, Y).$$

This means that $S : (x, y) \mapsto (T(x, y), y)$ is a Monge map that transports the independent coupling $P_X \otimes P_Y$ into the joint distribution P_{XY}:

$$S_{\#}(P_X \otimes P_Y) = P_{XY}.$$

(ii) *For every map T as in (i),*

$$P(T(X, y) \in \cdot) = P(X \in \cdot \,|Y = y), \quad dP_Y\text{-almost all } y \in F.$$

Conditional expectations as transports (Hosseini and Taghvaei 2022)

- Let $E = F = \mathbb{R}^d$ and $\mathcal{S}(P_X \otimes P_Y, P_{XY})$ be set of maps $S : (x, y) \mapsto (T(x, y), y)$ as above, and consider the transport problem over those maps:

$$\min_{S \in \mathcal{S}(P_X \otimes P_Y, P_{XY})} \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y} \left[\|T(X, Y) - X\|^2 \right].$$

Conditional expectations as transports (Hosseini and Taghvaei 2022)

- Let $E = F = \mathbb{R}^d$ and $\mathcal{S}(P_X \otimes P_Y, P_{XY})$ be set of maps $S : (x, y) \mapsto (T(x, y), y)$ as above, and consider the transport problem over those maps:

$$\min_{S \in \mathcal{S}(P_X \otimes P_Y, P_{XY})} \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y} \left[\|T(X, Y) - X\|^2 \right].$$

- Its dual reads as

$$\min_{f \in CVX_X} \mathbb{E}_{P_X \otimes P_Y} \left[f(X, Y) \right] + \mathbb{E}_{P_{XY}} \left[f^*(X, Y) \right],$$

where $f \in CVX_X$ iff $x \mapsto f(x, y)$ convex and in $L^1(P_X)$ for any y, and where $f^*(x, y) = \sup_z z \cdot x - f(z, y)$ is the convex conjugate of $f(\cdot, y)$.

Conditional expectations as transports (Hosseini and Taghvaei 2022)

- Let $E = F = \mathbb{R}^d$ and $\mathcal{S}(P_X \otimes P_Y, P_{XY})$ be set of maps $S : (x, y) \mapsto (T(x, y), y)$ as above, and consider the transport problem over those maps:

$$\min_{S \in \mathcal{S}(P_X \otimes P_Y, P_{XY})} \mathbb{E}_{(X,Y)\sim P_X \otimes P_Y} \left[\| T(X, Y) - X \|^2 \right].$$

- Its dual reads as

$$\min_{f \in CVX_X} \mathbb{E}_{P_X \otimes P_Y} \left[f(X, Y) \right] + \mathbb{E}_{P_{XY}} \left[f^*(X, Y) \right],$$

where $f \in CVX_X$ iff $x \mapsto f(x, y)$ convex and in $L^1(P_X)$ for any y, and where $f^*(x, y) = \sup_z z \cdot x - f(z, y)$ is the convex conjugate of $f(\cdot, y)$.

- Relation between the primal optimizer $\bar{T}$ and any dual optimizer $\bar{f}$:

$$\bar{T}(., y) = \nabla_x \bar{f}(., y),$$

so that

$$P_{X|Y=y} = \nabla_x \bar{f}(., y)_{\#} P_X$$

Example: Gaussian case

- Recall the Gaussian example $(X, Y) \sim \mathcal{N}(\mu, \Sigma)$, where for simplicity $\mu_i = 0, \sigma_i = 1$. Then we have
$$X = \rho Y + \sqrt{1-\rho^2}\gamma', \quad \gamma' \sim \mathcal{N}(0,1) \perp Y$$
- We can recover this by solving the OT problem above, that admits optimal transport map
$$\bar{T}(x, y) = \rho x + \sqrt{1-\rho^2} y$$
so that
$$P_{X|Y=y} = \bar{T}(., y)_{\#} P_X$$

The general (non-linear non-Gaussian) case

We want to develop an analogous analysis for systems of the form:

$$\begin{aligned} X_t &= g_t(X_{t-1}, \varepsilon_t), \quad X_0 \sim p_0 \\ Y_t &= h_t(X_t, \eta_t) \end{aligned}$$

The general (non-linear non-Gaussian) case

We want to develop an analogous analysis for systems of the form:

$$
\begin{aligned}
X_t &= g_t(X_{t-1}, \varepsilon_t), \quad X_0 \sim p_0 \\
Y_t &= h_t(X_t, \eta_t)
\end{aligned}
$$

- **I. Smoothing:** at every t, re-estimate all $\hat{X}_0, \hat{X}_1, ..., \hat{X}_t$, given $Y_0, ..., Y_t$.

The general (non-linear non-Gaussian) case

We want to develop an analogous analysis for systems of the form:

$$\begin{aligned} X_t &= g_t(X_{t-1}, \varepsilon_t), \qquad X_0 \sim p_0 \\ Y_t &= h_t(X_t, \eta_t) \end{aligned}$$

- **I. Smoothing:** at every t, re-estimate all $\hat{X}_0, \hat{X}_1, ..., \hat{X}_t$, given $Y_0, ..., Y_t$.
- **II. Non-smoothing:** at every t, keep previous estimates $\hat{X}_0, \hat{X}_1, ..., \hat{X}_{t-1}$, and estimate only $\hat{X}_t$ using:
  - previous estimates, together with
  - new observation Y_t

Smoothing

I. Smoothing: at every t, re-estimate all $\hat{X}_0, \hat{X}_1, ..., \hat{X}_t$, given $Y_0, ..., Y_t$.

Smoothing

I. Smoothing: at every t, re-estimate all $\hat{X}_0, \hat{X}_1, ..., \hat{X}_t$, given $Y_0, ..., Y_t$.

- Consider $T_t : \mathbb{R}^{2d(t+1)} \longrightarrow \mathbb{R}^{d(t+1)}$ and $S_t : \mathbb{R}^{2d(t+1)} \longrightarrow \mathbb{R}^{2d(t+1)}$, $S_t(x, y) = (T_t(x, y), y)$ s.t.

$$S_{t\#}(P_{X_{0:t}} \otimes P_{Y_{0:t}}) = P_{X_{0:t}, Y_{0:t}},$$

so that $T_t(X_{0:t}; Y_{0:t})$ has the interpretation of $X_{0:t}|Y_{0:t}$

Smoothing

I. Smoothing: at every t, re-estimate all $\hat{X}_0, \hat{X}_1, ..., \hat{X}_t$, given $Y_0, ..., Y_t$.

- Consider $T_t : \mathbb{R}^{2d(t+1)} \to \mathbb{R}^{d(t+1)}$ and $S_t : \mathbb{R}^{2d(t+1)} \to \mathbb{R}^{2d(t+1)}$, $S_t(x,y) = (T_t(x,y), y)$ s.t.
$$S_{t\#}(P_{X_{0:t}} \otimes P_{Y_{0:t}}) = P_{X_{0:t},Y_{0:t}},$$
so that $T_t(X_{0:t}; Y_{0:t})$ has the interpretation of $X_{0:t}|Y_{0:t}$
- Consider the transport problem with cost $||T_t(X_{0:t}; Y_{0:t}) - X_{0:t}||^2$ over such maps S_t

$\Rightarrow$ $t+1$-dimensional version of the static setting seen above: solve dual problem and get $\bar{f}$, and from it obtain, for any observation $y_{0:t}$:

$$P_{X_{0:t}|Y_{0:t}=y_{0:t}} = \nabla_x \bar{f}(., y_{0:t})_\# P_{X_{0:t}}$$

Smoothing - algorithm

- At time t we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_{0:t}} \otimes P_{Y_{0:t}}} \left[f(X, Y) \right] + \mathbb{E}_{P_{X_{0:t}, Y_{0:t}}} \left[f^*(X, Y) \right]$$

Smoothing - algorithm

- At time t we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_{0:t}} \otimes P_{Y_{0:t}}} \left[f(X,Y) \right] + \mathbb{E}_{P_{X_{0:t},Y_{0:t}}} \left[f^*(X,Y) \right]$$

- Sample $\{X^i_{0:t}\}_{i=1,..,N}$ from prior $P_{X_{0:t}}$ and from them generate $Y^i_{0:t} \sim P_{Y_{0:t}|X_{0:t}=X^i_{0:t}}$ so that $\{(X^i_{0:t}, Y^i_{0:t})\}_{i=1,..,N}$ is an independent sample from the joint distribution $P_{X_{0:t},Y_{0:t}}$

Smoothing - algorithm

- At time t we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_{0:t}} \otimes P_{Y_{0:t}}} \left[f(X,Y) \right] + \mathbb{E}_{P_{X_{0:t},Y_{0:t}}} \left[f^*(X,Y) \right]$$

- Sample $\{X^i_{0:t}\}_{i=1,..,N}$ from prior $P_{X_{0:t}}$ and from them generate $Y^i_{0:t} \sim P_{Y_{0:t}|X_{0:t}=X^i_{0:t}}$ so that $\{(X^i_{0:t}, Y^i_{0:t})\}_{i=1,..,N}$ is an independent sample from the joint distribution $P_{X_{0:t},Y_{0:t}}$
- Fix a subset $\mathcal{F} \subset CVX_X$ of parameterized functions and define the empirical cost

$$V^N(f) = \frac{1}{N(N-1)} \sum_{i \neq j=1}^{N} f(X^i_{0:t}, Y^j_{0:t}) + \frac{1}{N} \sum_{i=1}^{N} f^*(X^i_{0:t}, Y^i_{0:t}), \quad \forall\, f \in \mathcal{F}$$

Smoothing - algorithm

- At time t we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_{0:t}} \otimes P_{Y_{0:t}}} \left[f(X,Y) \right] + \mathbb{E}_{P_{X_{0:t},Y_{0:t}}} \left[f^*(X,Y) \right]$$

- Sample $\{X_{0:t}^i\}_{i=1,..,N}$ from prior $P_{X_{0:t}}$ and from them generate $Y_{0:t}^i \sim P_{Y_{0:t}|X_{0:t}=X_{0:t}^i}$ so that $\{(X_{0:t}^i, Y_{0:t}^i)\}_{i=1,..,N}$ is an independent sample from the joint distribution $P_{X_{0:t},Y_{0:t}}$
- Fix a subset $\mathcal{F} \subset CVX_X$ of parameterized functions and define the empirical cost

$$V^N(f) = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N f(X_{0:t}^i, Y_{0:t}^j) + \frac{1}{N} \sum_{i=1}^N f^*(X_{0:t}^i, Y_{0:t}^i), \quad \forall f \in \mathcal{F}$$

- Minimize over $\mathcal{F}$ and use $\bar{f}^{N,\mathcal{F}} \in \underset{f \in \mathcal{F}}{\operatorname{argmin}} V^N(f)$ to generate sample from posterior given the realization $y_{0:t}$:

$$\underset{\text{posterior}}{\underset{\uparrow}{\tilde{X}_{0:t}^i}} = \nabla_x \bar{f}^{N,\mathcal{F}} (\underset{\text{prior}}{\underset{\uparrow}{X_{0:t}^i}}, \underset{\text{observation}}{\underset{\uparrow}{y_{0:t}}})$$

Non-smoothing

II. Non-smoothing: at every t, keep previous estimates $\hat{X}_0, \hat{X}_1, ..., \hat{X}_{t-1}$, and estimate $\hat{X}_t$ using the previous estimates together with the new observation Y_t

Non-smoothing

II. Non-smoothing: at every t, keep previous estimates $\hat{X}_0, \hat{X}_1, ..., \hat{X}_{t-1}$, and estimate $\hat{X}_t$ using the previous estimates together with the new observation Y_t

Idea: use the two-step iteration

$$\begin{aligned}\hat{X}_t^- &= \mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)] \\ \hat{X}_t &= \textit{function}(\hat{X}_t^-, Y_t)\end{aligned}$$

learning the conditioning function as an *optimal transport map*, analogous to HT22

Non-smoothing

II. Non-smoothing: at every t, keep previous estimates $\hat{X}_0, \hat{X}_1, ..., \hat{X}_{t-1}$, and estimate $\hat{X}_t$ using the previous estimates together with the new observation Y_t

Idea: use the two-step iteration

$$\begin{aligned} \hat{X}_t^- &= \mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)] \\ \hat{X}_t &= \textit{function}(\hat{X}_t^-, Y_t) \end{aligned}$$

learning the conditioning function as an *optimal transport map*, analogous to HT22

“Something like”

$$\min_{S_t \in \mathcal{S}(P_{X_t} \otimes P_{Y_t}, P_{X_t, Y_t})} \mathbb{E}_{(X_t, Y_t) \sim P_{X_t} \otimes P_{Y_t}} \left[\|T_t(X_t; Y_t) - X_t\|^2 \right]$$

Non-smoothing

II. Non-smoothing: at every t, keep previous estimates $\hat{X}_0, \hat{X}_1, ..., \hat{X}_{t-1}$, and estimate $\hat{X}_t$ using the previous estimates together with the new observation Y_t

Idea: use the two-step iteration

$$\begin{aligned} \hat{X}_t^- &= \mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)] \\ \hat{X}_t &= \textit{function}(\hat{X}_t^-, Y_t) \end{aligned}$$

learning the conditioning function as an *optimal transport map*, analogous to HT22

"Something like"

$$\min_{S_t \in \mathcal{S}(P_{X_t} \otimes P_{Y_t}, P_{X_t, Y_t})} \mathbb{E}_{(X_t, Y_t) \sim P_{X_t} \otimes P_{Y_t}} \left[\|T_t(X_t; Y_t) - X_t\|^2 \right]$$

→ But some adjustment is needed since $\hat{X}_t^-$ and Y_t are NOT independent

Non-smoothing

- We want to set $\hat{X}_t = \textit{function}(\mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)], Y_t)$ with independent arguments

Non-smoothing

- We want to set $\hat{X}_t = \mathit{function}(\mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)], Y_t)$ with independent arguments
  $\Rightarrow$ condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$

Non-smoothing

- We want to set $\hat{X}_t = \textit{function}(\mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)], Y_t)$ with independent arguments
  $\Rightarrow$ condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$
- Let the map $\bar{T}_t^{\bar{x}}$ be s.t. $S_t(x,y) = (\bar{T}_t^{\bar{x}}(x,y), y)$ is optimizer for

$$\min_{S_t \in \mathcal{S}\left(P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}, P_{(X_t,Y_t)|X_{t-1}=\bar{x}}\right)} \mathbb{E}_{(X_t,Y_t)\sim P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}} \left[\|T_t(X_t; Y_t) - X_t\|^2\right],$$

- i.e. $\bar{T}_t^{\bar{x}}(.,y) = \nabla_x \bar{f}_t^{\bar{x}}(.,y)$, with $\bar{f}_t^{\bar{x}}$ dual optimizer

Non-smoothing

- We want to set $\hat{X}_t = \textit{function}(\mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\hat{X}_{t-1}, \tilde{\varepsilon}_t)], Y_t)$ with independent arguments
  $\Rightarrow$ condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$
- Let the map $\bar{T}_t^{\bar{x}}$ be s.t. $S_t(x, y) = (\bar{T}_t^{\bar{x}}(x, y), y)$ is optimizer for

$$\min_{S_t \in \mathcal{S}\left(P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}, P_{(X_t,Y_t)|X_{t-1}=\bar{x}}\right)} \mathbb{E}_{(X_t,Y_t) \sim P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}} \left[\|T_t(X_t; Y_t) - X_t\|^2\right],$$

- i.e. $\bar{T}_t^{\bar{x}}(., y) = \nabla_x \bar{f}_t^{\bar{x}}(., y)$, with $\bar{f}_t^{\bar{x}}$ dual optimizer
- As updating step in our algorithm, we take

$$\hat{X}_t = \nabla_x \bar{f}_t^{\bar{x}}(\mathbb{E}_{\tilde{\varepsilon}_t \sim \varepsilon_t}[g_t(\bar{x}, \tilde{\varepsilon}_t)], Y_t)$$

Example: Kalman filter

- Recall the system

$$X_t = a_t X_{t-1} + b_t \varepsilon_t,$$
$$Y_t = A_t X_t + B_t \eta_t,$$

with ε_t, η_t independent standard normal, where we have

$$\hat{X}_t = a_t \hat{X}_{t-1} + G_t \cdot (Y_t - A_t a_t \hat{X}_{t-1})$$

- We can recover this by solving the OT problems above, that admit optimal transport map (same for every $\bar{x}$)

$$\bar{T}_t(x; y) = x + G_t \cdot (y - A_t x)$$

Non-smoothing - algorithm

- At time t, condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$, we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}} \left[f(X,Y) \right] + \mathbb{E}_{P_{(X_t,Y_t)|X_{t-1}=\bar{x}}} \left[f^*(X,Y) \right]$$

Non-smoothing - algorithm

- At time t, condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$, we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}} \left[f(X, Y) \right] + \mathbb{E}_{P_{(X_t,Y_t)|X_{t-1}=\bar{x}}} \left[f^*(X, Y) \right]$$

- Sample $\{\tilde{\varepsilon}_t^i\}_{i=1,..,N} \sim \varepsilon_t$ independent from everything else, to get the sample $X_t^i = g_t(\bar{x}, \tilde{\varepsilon}_t^i)$ from the prior and from them generate $Y_t^i \sim P_{Y_t|X_t=X_t^i}$, so that $\{(X_t^i, Y_t^i)\}_{i=1,..,N}$ is an independent sample from the joint distribution $P_{(X_t,Y_t)|X_{t-1}=\bar{x}}$

Non-smoothing - algorithm

- At time t, condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$, we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}} \left[f(X,Y) \right] + \mathbb{E}_{P_{(X_t,Y_t)|X_{t-1}=\bar{x}}} \left[f^*(X,Y) \right]$$

- Sample $\{\tilde{\varepsilon}_t^i\}_{i=1,..,N} \sim \varepsilon_t$ independent from everything else, to get the sample $X_t^i = g_t(\bar{x}, \tilde{\varepsilon}_t^i)$ from the prior and from them generate $Y_t^i \sim P_{Y_t|X_t=X_t^i}$, so that $\{(X_t^i, Y_t^i)\}_{i=1,..,N}$ is an independent sample from the joint distribution $P_{(X_t,Y_t)|X_{t-1}=\bar{x}}$
- Fix a subset $\mathcal{F} \subset CVX_X$ of parameterized functions and define the empirical cost

$$V^N(f) = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N f(X_t^i, Y_t^j) + \frac{1}{N} \sum_{i=1}^N f^*(X_t^i, Y_t^i), \quad \forall f \in \mathcal{F}$$

Non-smoothing - algorithm

- At time t, condition on the previous estimate $\hat{X}_{t-1} = \bar{x}$, we face the dual problem:

$$\min_{f \in CVX_X} \mathbb{E}_{P_{X_t|X_{t-1}=\bar{x}} \otimes P_{Y_t|X_{t-1}=\bar{x}}} \left[f(X, Y) \right] + \mathbb{E}_{P_{(X_t,Y_t)|X_{t-1}=\bar{x}}} \left[f^*(X, Y) \right]$$

- Sample $\{\tilde{\varepsilon}_t^i\}_{i=1,..,N} \sim \varepsilon_t$ independent from everything else, to get the sample $X_t^i = g_t(\bar{x}, \tilde{\varepsilon}_t^i)$ from the prior and from them generate $Y_t^i \sim P_{Y_t|X_t=X_t^i}$, so that $\{(X_t^i, Y_t^i)\}_{i=1,..,N}$ is an independent sample from the joint distribution $P_{(X_t,Y_t)|X_{t-1}=\bar{x}}$
- Fix a subset $\mathcal{F} \subset CVX_X$ of parameterized functions and define the empirical cost

$$V^N(f) = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N f(X_t^i, Y_t^j) + \frac{1}{N} \sum_{i=1}^N f^*(X_t^i, Y_t^i), \quad \forall f \in \mathcal{F}$$

- Minimize over $\mathcal{F}$ and use $f^{\bar{x},N,\mathcal{F}} \in \underset{f \in \mathcal{F}}{\operatorname{argmin}} V^N(f)$ to generate sample from posterior given the realization y_t:

$$\tilde{X}_t^i = \nabla_x f^{\bar{x},N,\mathcal{F}}(X_t^i, y_t)$$

Conclusions

- We consider discrete-time dynamic systems of hidden and observable processes (these can be defined on different spaces, have different dimension etc)

Conclusions

- We consider discrete-time dynamic systems of hidden and observable processes (these can be defined on different spaces, have different dimension etc)
- We consider smoothing and non-smoothing in filtering

Conclusions

- We consider discrete-time dynamic systems of hidden and observable processes (these can be defined on different spaces, have different dimension etc)
- We consider smoothing and non-smoothing in filtering
- We implement a variational representation of Bayes' through optimal transport theory, to learn transport maps that push independent coupling to joint distribution, and so implicitly the map that sends prior to posterior (we are not only learning $\hat{X}_t = \mathbb{E}[X_t|Y_0, ..., Y_t]$ but the whole posterior distribution $P_{X_t|Y_0,...,Y_t}$)

Conclusions

- We consider discrete-time dynamic systems of hidden and observable processes (these can be defined on different spaces, have different dimension etc)
- We consider smoothing and non-smoothing in filtering
- We implement a variational representation of Bayes' through optimal transport theory, to learn transport maps that push independent coupling to joint distribution, and so implicitly the map that sends prior to posterior (we are not only learning $\hat{X}_t = \mathbb{E}[X_t | Y_0, ..., Y_t]$ but the whole posterior distribution $P_{X_t | Y_0, ..., Y_t}$)
- Once optimal transport maps are learned (by simulation and approximation of dual problem), these can be used for any realization of the observable process (without need to be computed again for different realizations)

Still a lot to do and understand...

- Reduce complexity by e.g. solving some conditional version of the transport problem, or by using $\tilde{X}_{t-1}^{i}$ rather than conditioning to $\hat{X}_{t-1} = \bar{x}$

Still a lot to do and understand...

- Reduce complexity by e.g. solving some conditional version of the transport problem, or by using $\tilde{X}^i_{t-1}$ rather than conditioning to $\hat{X}_{t-1} = \bar{x}$
- Consider more general dynamics

Still a lot to do and understand...

- Reduce complexity by e.g. solving some conditional version of the transport problem, or by using $\tilde{X}^i_{t-1}$ rather than conditioning to $\hat{X}_{t-1} = \bar{x}$
- Consider more general dynamics
- Consider different costs in the OT problem, test and compare solutions

Still a lot to do and understand...

- Reduce complexity by e.g. solving some conditional version of the transport problem, or by using $\tilde{X}^i_{t-1}$ rather than conditioning to $\hat{X}_{t-1} = \bar{x}$
- Consider more general dynamics
- Consider different costs in the OT problem, test and compare solutions
- Study stability, error, ...

Still a lot to do and understand...

- Reduce complexity by e.g. solving some conditional version of the transport problem, or by using $\tilde{X}^i_{t-1}$ rather than conditioning to $\hat{X}_{t-1} = \bar{x}$
- Consider more general dynamics
- Consider different costs in the OT problem, test and compare solutions
- Study stability, error, ...
- Introduce uncertainty (around dynamics of X or Y, or as adapted Wasserstein balls around P_X, P_Y or P_{XY})

Still a lot to do and understand...

- Reduce complexity by e.g. solving some conditional version of the transport problem, or by using $\tilde{X}^i_{t-1}$ rather than conditioning to $\hat{X}_{t-1} = \bar{x}$
- Consider more general dynamics
- Consider different costs in the OT problem, test and compare solutions
- Study stability, error, ...
- Introduce uncertainty (around dynamics of X or Y, or as adapted Wasserstein balls around P_X, P_Y or P_{XY})

Thank you for your attention!