Neural Newtork approximation of the Mortensen observer in high-dimension

Martin Chassard[1], Virginie Ehrlacher[12], Damiano Lombardi[34], Philippe Moireau[56]

[1]CERMICS, ENPC [2]INRIA-MATHERIALS [3]LJLL, Sorbonne Université
[4]INRIA-COMMEDIA [5]CMAP, École Polytechnique [6]INRIA-ANANKE

Tuesday 2nd June, CANUM 2026

We consider the **nonlinear deterministic filtering** problem (Fleming 1997) in **compact** state space $\mathcal{X} \subset \mathbb{R}^d$ and observation space $\mathcal{Y} \subset \mathbb{R}^m$, $m \leq d$.

$$\begin{cases} \dot{x}(t) = f(x(t)) + \nu(t), & x(0) = \hat{\mathrm{x}}_0 + \zeta, \\ y(t) = h(x(t)) + \eta(t), & t \in (0, T). \end{cases} \tag{1}$$

Assumption

- $\nu \in L^2((0,T), \mathcal{X}) =: \mathcal{V}_T$, $\eta \in L^2((0,T); \mathcal{Y})$;
- $f, h \in C^1, \nabla f, \nabla h$ *bounded and* f, h *sub-linear:*

$$\|f(x)\| \leq C_f(1 + \|x\|), \quad \|h(x)\| \leq C_h(1 + \|x\|). \tag{2}$$

We consider the **nonlinear deterministic filtering** problem (Fleming 1997) in **compact** state space $\mathcal{X} \subset \mathbb{R}^d$ and observation space $\mathcal{Y} \subset \mathbb{R}^m$, $m \leq d$.

$$\begin{cases} \dot{x}(t) = f(x(t)) + \nu(t), & x(0) = \hat{\mathrm{x}}_0 + \zeta, \\ y(t) = h(x(t)) + \eta(t), & t \in (0,T). \end{cases} \tag{1}$$

Assumption

- $\nu \in L^2((0,T),\mathcal{X}) =: \mathcal{V}_T$, $\eta \in L^2((0,T);\mathcal{Y})$;
- $f, h \in C^1, \nabla f, \nabla h$ *bounded and* f, h *sub-linear:*

$$\|f(x)\| \leq C_f(1+\|x\|), \quad \|h(x)\| \leq C_h(1+\|x\|). \tag{2}$$

Under the assumptions, the dynamic system is well-posed $\implies$ we denote $x|_{\zeta,\nu}$ the trajectory associated to $(\zeta,\nu) \in \mathcal{X} \times \mathcal{V}_T$.

The Mortensen observer

Optimal estimation

Let $t \in (0,T)$, $\Lambda_0, S \in \mathbb{S}^d_{++}, R \in \mathbb{S}^m_{++}$, we consider minimizing

$$\mathscr{J}_t(\zeta,\nu) := \frac{1}{2}\|\zeta\|^2_{\Lambda_0} + \frac{1}{2}\int_0^t \left[\|\nu(s)\|^2_S + \|y(s) - h(x|_{\zeta,\nu}(s))\|^2_R\right] \mathrm{d}s. \tag{3}$$

Proposition

Let $t \in (0,T)$, there exists at least one minimizer

$$(\overline{\zeta}_t, \overline{\nu}_t) \in \underset{(\zeta,\nu)\in\mathcal{X}\times\mathcal{V}_t}{\arg\min}\, \mathscr{J}_t(\zeta,\nu).$$

Optimal estimation

Let $t \in (0,T)$, $\Lambda_0, S \in \mathbb{S}^d_{++}$, $R \in \mathbb{S}^m_{++}$, we consider minimizing

$$\mathscr{J}_t(\zeta,\nu) := \frac{1}{2}\|\zeta\|^2_{\Lambda_0} + \frac{1}{2}\int_0^t \left[\|\nu(s)\|^2_S + \|y(s) - h(x|_{\zeta,\nu}(s))\|^2_R\right] \mathrm{d}s. \tag{3}$$

Proposition

Let $t \in (0,T)$, there exists at least one minimizer

$$(\overline{\zeta}_t, \overline{\nu}_t) \in \underset{(\zeta,\nu)\in\mathcal{X}\times\mathcal{V}_t}{\arg\min}\ \mathscr{J}_t(\zeta,\nu).$$

$\Longrightarrow$ minimizing trajectory $\overline{x}_t := x|_{\overline{\zeta}_t,\overline{\nu}_t}$ and instantaneous least-squares estimator

$$\hat{\mathrm{x}}(t) := \overline{x}_t(t). \tag{4}$$

The estimator is also called **Mortensen observer** (Mortensen 1968).

Definition (Observer)

Assume $T = +\infty$. A time function $z : (0,T) \to \mathcal{X}$ is said to be an observer for the dynamics if $\|z(t) - x(t)\| \to 0$, as $t \to +\infty$.

The value function and HJB

Let $(t,x) \in (0,T) \times \mathcal{X}$. We define the value function associated with the control problem

$$V(t,x) = \begin{cases} \min\limits_{(\zeta,\nu)\in\mathcal{X}\times\mathcal{V}_t} \mathscr{J}_t(\zeta,\nu), \\ \text{s.t.} \quad x|_{\zeta,\nu}(t) = x. \end{cases} \tag{5}$$

Proposition (HJB)

The value function is the unique viscosity solution (Crandall and Lions 1983) to the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t V(t,x) + \frac{1}{2}\langle f(x), \nabla V(t,x)\rangle + \frac{1}{2}\|\nabla V(t,x)\|_S^2 \\ \qquad -\frac{1}{2}\|y(t) - h(t,x)\|_R^2 = 0, \\ V(0,x) = \frac{1}{2}\|x - \hat{\mathrm{x}}_0\|_{\Lambda_0}^2. \end{cases} \tag{HJB}$$

The value function and HJB

Let $(t, x) \in (0, T) \times \mathcal{X}$. We define the value function associated with the control problem

$$V(t,x) = \begin{cases} \min\limits_{(\zeta,\nu)\in\mathcal{X}\times\mathcal{V}_t} \mathscr{J}_t(\zeta,\nu), \\ \text{s.t.} \quad x|_{\zeta,\nu}(t) = x. \end{cases} \tag{5}$$

Proposition (HJB)

The value function is the unique viscosity solution (Crandall and Lions 1983) to the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t V(t,x) + \frac{1}{2}\langle f(x), \nabla V(t,x)\rangle + \frac{1}{2}\|\nabla V(t,x)\|_S^2 \\ \qquad -\frac{1}{2}\|y(t) - h(t,x)\|_R^2 = 0, \\ V(0,x) = \frac{1}{2}\|x - \hat{\mathrm{x}}_0\|_{\Lambda_0}^2. \end{cases} \tag{HJB}$$

Proposition (HJB-based estimation)

The estimator satisfies for all $t \in (0, T)$,

$$\hat{\mathrm{x}}(t) \in \arg\min_{x\in\mathcal{X}} V(t,x). \tag{6}$$

The LQ case

Assume linearity:

$$f : x \mapsto Ax, \quad h : x \mapsto Cx,$$

for matrices $A \in \mathbb{R}^{d\times d}, C \in \mathbb{R}^{m\times d}$.

Then the value function is explicitly given by

$$V(t,x) = \frac{1}{2}\langle x - \hat{x}(t), \Pi^{-1}(t)(x - \hat{x}(t))\rangle + r(t), \tag{7}$$

where $(\hat{x}, \Pi)$ are obtained by the continuous Kalman filter (Kalman and Bucy 1961):

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + \Pi C^\top R(y - C\hat{x}), \\ \dot{\Pi} = A\Pi + \Pi A^\top - \Pi C^\top R C \Pi + S^{-1}. \end{cases} \tag{8}$$

The LQ case

Assume linearity:

$$f : x \mapsto Ax, \quad h : x \mapsto Cx,$$

for matrices $A \in \mathbb{R}^{d\times d}, C \in \mathbb{R}^{m\times d}$.

Then the value function is explicitly given by

$$V(t,x) = \frac{1}{2}\langle x - \hat{x}(t), \Pi^{-1}(t)(x - \hat{x}(t))\rangle + r(t), \tag{7}$$

where $(\hat{x}, \Pi)$ are obtained by the continuous Kalman filter (Kalman and Bucy 1961):

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + \Pi C^\top R(y - C\hat{x}), \\ \dot{\Pi} = A\Pi + \Pi A^\top - \Pi C^\top R C \Pi + S^{-1}. \end{cases} \tag{8}$$

We are concerned with the numerical approximation of the Mortensen observer based on HJB in the high-dimensional ($d \gg 1$) setting.

Remarks

1. (HJB) is a non-linear equation exhibiting singularities (convexity not conserved along the flow, creation of cusps);

We are concerned with the numerical approximation of the Mortensen observer based on HJB in the high-dimensional ($d \gg 1$) setting.

Remarks

1. (HJB) is a non-linear equation exhibiting singularities (convexity not conserved along the flow, creation of cusps);
2. we are suffering from the **curse of dimensionality** $\Longrightarrow$ grid-based methods scale exponentially with d;

We are concerned with the numerical approximation of the Mortensen observer based on HJB in the high-dimensional ($d \gg 1$) setting.

Remarks

1. (HJB) is a non-linear equation exhibiting singularities (convexity not conserved along the flow, creation of cusps);
2. we are suffering from the **curse of dimensionality** $\implies$ grid-based methods scale exponentially with d;
3. to avoid solving the high-dimensional HJB equation, what is usually done is to linearize the system via the so-called **Extended Kalman Filter** (EKF).

Outline

Time discretization

Let us consider a discrete version of the dynamics/observation system

$$\begin{cases} x_{n+1} = \varphi^{\tau}_{n|n+1}(x_n) + \nu_n, \\ y_n = h(x_n) + \eta_n, \quad n \in \mathbb{N}. \end{cases} \tag{9}$$

for some time discretization $\tau > 0$.

Time discretization

Let us consider a discrete version of the dynamics/observation system

$$\begin{cases} x_{n+1} = \varphi^{\tau}_{n|n+1}(x_n) + \nu_n, \\ y_n = h(x_n) + \eta_n, \quad n \in \mathbb{N}. \end{cases} \tag{9}$$

for some time discretization $\tau > 0$.

Time splitting scheme

We will be considering the scheme given for all $n \in \mathbb{N}$ by:

$$\begin{cases} V_{0^-}(x) = \frac{1}{2}\|x - \hat{\mathrm{x}}_0\|^2_{\Lambda_0}, \quad \text{(initialization)}; \\ V_{n^+}(x) = V_{n^-}(x) + \frac{\tau}{2}\|y_n - h(x)\|^2_R, \quad \text{(correction)}; \\ V_{n+1^-}(x) = (V_{n^+} \circ (\varphi^{\tau}_{n|n+1})^{-1}) \,\square \left(\frac{1}{2\tau}\|\cdot\|^2_S\right)(x), \quad \text{(prediction)}; \end{cases}$$

where the inf-convolution is defined as

$$f \,\square\, g(x) = \inf_{z \in \mathcal{X}} \{f(z) + g(z - x)\}. \tag{10}$$

One can show that the scheme **converges** as $\tau \to 0$ (consistency : Moireau 2018).

The LQ case in a discrete setting

Assuming linearity, we write $\varphi^{\tau}_{n|n+1}(x) = \Phi_{n|n+1}x$, for some matrix $\Phi_{n|n+1} \in \mathbb{R}^{d\times d}$.
Then the discrete value function writes for every $n \in \mathbb{N}$

$$V_{n^\pm}(x) = \frac{1}{2}\langle x - \hat{x}_{n^\pm}, \Pi^{-1}_{n^\pm}(x - \hat{x}_{n^\pm})\rangle + r_{n^\pm}, \qquad (11)$$

where $(\hat{x}_{n^\pm}, \Pi_{n^\pm})$ satisfy the equations of the discrete Kalman filter.

The LQ case in a discrete setting

Assuming linearity, we write $\varphi^{\tau}_{n|n+1}(x) = \Phi_{n|n+1}x$, for some matrix $\Phi_{n|n+1} \in \mathbb{R}^{d\times d}$.
Then the discrete value function writes for every $n \in \mathbb{N}$

$$V_{n\pm}(x) = \frac{1}{2}\langle x - \hat{x}_{n\pm}, \Pi^{-1}_{n\pm}(x - \hat{x}_{n\pm})\rangle + r_{n\pm}, \tag{11}$$

where $(\hat{x}_{n\pm}, \Pi_{n\pm})$ satisfy the equations of the discrete Kalman filter.

The Discrete Kalman Filter

It is given by the following time-splitting scheme:

$$\forall n \in \mathbb{N}, \begin{cases} \hat{x}_{n+} = \hat{x}_{n-} + \Pi_{n+}C_n^{\top}R_n(y_n - C_n\hat{x}_{n-}), & \text{(correction)} \\ \hat{x}_{n+1-} = \Phi_{n|n+1}\hat{x}_{n+}; & \text{(prediction)} \end{cases} \tag{12}$$

$$\forall n \in \mathbb{N}, \begin{cases} \Pi_{n+} = (\mathrm{Id} - G_nC_n)\Pi_{n-}(\mathrm{Id} - G_nC_n)^{\top} + G_nR_n^{-1}G_n^{\top}, \\ G_n := \Pi_{n-}C_n^{\top}(R_n^{-1} + C_n\Pi_{n-}C_n^{\top})^{-1}, & \text{(correction)} \\ \Pi_{n+1-} = \Phi_{n|n+1}\Pi_{n+}\Phi_{n|n+1}^{\top} + S_n^{-1}. & \text{(prediction)} \end{cases} \tag{13}$$

Outline

Problem setting

Time discretization of HJB

Neural Network representations and Optimal Transport : EKF-OT

Numerical illustrations

We will be using tools from Optimal Transport.

The Monge OT problem

Let $\mu, \mathfrak{m} \in \mathscr{P}(\mathcal{X})$, the Monge formulation of the (Euclidian) OT problem writes

$$\inf_{\mathfrak{t}_\# \mu = \mathfrak{m}} \int_{\mathcal{X}} \|x - \mathfrak{t}(x)\|^2 \, \mathrm{d}\mu(x). \tag{14}$$

Theorem (Brenier)

Let $\mu, \mathfrak{m} \in \mathscr{P}_{\mathrm{ac}}(\mathcal{X})$ be a.c. probability measures. Then, there exists a convex function $\phi_{\mu\to\mathfrak{m}}$, called Kantorovich potential, such that $\mathfrak{t} = \nabla\phi_{\mu\to\mathfrak{m}}$ is solution to the Monge OT problem.

We will be using tools from Optimal Transport.

The Monge OT problem

Let $\mu, \mathfrak{m} \in \mathscr{P}(\mathcal{X})$, the Monge formulation of the (Euclidian) OT problem writes

$$\inf_{\mathfrak{t}_{\#}\mu=\mathfrak{m}} \int_{\mathcal{X}} \|x - \mathfrak{t}(x)\|^2 \, \mathrm{d}\mu(x). \tag{14}$$

Theorem (Brenier)

Let $\mu, \mathfrak{m} \in \mathscr{P}_{\mathrm{ac}}(\mathcal{X})$ be a.c. probability measures. Then, there exists a convex function $\phi_{\mu\to\mathfrak{m}}$, called Kantorovich potential, such that $\mathfrak{t} = \nabla\phi_{\mu\to\mathfrak{m}}$ is solution to the Monge OT problem.

Proposition (Perturbation of gaussians)

Let $\mu \sim \mathcal{N}(m, \Pi)$ be a gaussian measure and let $\mathfrak{m} \in \mathscr{P}_{\mathrm{ac}}(\mathcal{X})$ be an a.c. probability measure. Then, the measure $\mathfrak{m}$ can be written

$$\mathfrak{m} \propto \exp\left(-\frac{1}{2}\langle \nabla\phi^*_{\mu\to\mathfrak{m}}(x) - m, \Pi^{-1}(\nabla\phi^*_{\mu\to\mathfrak{m}}(x) - m)\rangle\right). \tag{15}$$

$\Longrightarrow$ **main idea** : Learn Kantorovich potentials ϕ using Input Convex Neural Networks (ICNN, Amos, Xu, and Kolter 2017).

EKF-based OT algorithm

We rely on EKF and correct the nonlinearity with OT.

EKF-OT (correction step; prediciton step not shown but similar)

Let $n \in \mathbb{N}$, and assume we have computed approximations $(\hat{\mathrm{x}}_{n^-}, \Pi_{n^-})$.

1. Compute $(\hat{x}_{n^+}, \Pi_{n^+})$ by EKF;

EKF-based OT algorithm

We rely on EKF and correct the nonlinearity with OT.

EKF-OT (correction step; prediciton step not shown but similar)

Let $n \in \mathbb{N}$, and assume we have computed approximations $(\hat{\mathrm{x}}_{n-}, \Pi_{n-})$.

1. Compute $(\hat{x}_{n+}, \Pi_{n+})$ by EKF;
2. initialize an ICNN g_θ and define the ansatz

$$V_{n+|\theta}(x) = \frac{1}{2}\langle \nabla g_\theta(x) - \hat{x}_{n+}, \Pi_{n+}^{-1}(\nabla g_\theta(x) - \hat{x}_{n+})\rangle + \tilde{r}_{n+}; \quad (16)$$

We rely on EKF and correct the nonlinearity with OT.

EKF-OT (correction step; prediciton step not shown but similar)

Let $n \in \mathbb{N}$, and assume we have computed approximations $(\hat{\mathrm{x}}_{n-}, \Pi_{n-})$.

1. Compute $(\hat{x}_{n+}, \Pi_{n+})$ by EKF;
2. initialize an ICNN g_θ and define the ansatz

$$V_{n+|\theta}(x) = \frac{1}{2}\langle \nabla g_\theta(x) - \hat{x}_{n+}, \Pi_{n+}^{-1}(\nabla g_\theta(x) - \hat{x}_{n+})\rangle + \tilde{r}_{n+}; \quad (16)$$

3. for $k \in \{1, \cdots, K\}$, sample $X^{(k)} \sim \mathcal{N}(\hat{x}_{n+}, \Pi_{n+})$ and define $\mu_{n+}^{(K)} = \frac{1}{K}\sum_{k=1}^{K} \delta_{X^{(k)}}$;

EKF-based OT algorithm

We rely on EKF and correct the nonlinearity with OT.

EKF-OT (correction step; prediciton step not shown but similar)

Let $n \in \mathbb{N}$, and assume we have computed approximations $(\hat{\mathrm{x}}_{n-}, \Pi_{n-})$.

1. Compute $(\hat{x}_{n+}, \Pi_{n+})$ by EKF;
2. initialize an ICNN g_θ and define the ansatz

$$V_{n+|\theta}(x) = \frac{1}{2}\langle \nabla g_\theta(x) - \hat{x}_{n+}, \Pi_{n+}^{-1}(\nabla g_\theta(x) - \hat{x}_{n+})\rangle + \tilde{r}_{n+}; \quad (16)$$

3. for $k \in \{1, \cdots, K\}$, sample $X^{(k)} \sim \mathcal{N}(\hat{x}_{n+}, \Pi_{n+})$ and define $\mu_{n+}^{(K)} = \frac{1}{K}\sum_{k=1}^{K} \delta_{X^{(k)}}$;
4. define the least-squares loss function

$$L_{n+}(\theta) = \mathbb{E}_{X\sim\mu_{n+}^{(K)}}\left\{\left|V_{n+|\theta}(X) - V_{n+}(X)\right|^2\right\}; \quad (17)$$

EKF-based OT algorithm

We rely on EKF and correct the nonlinearity with OT.

EKF-OT (correction step; prediciton step not shown but similar)

Let $n \in \mathbb{N}$, and assume we have computed approximations $(\hat{\mathrm{x}}_{n-}, \Pi_{n-})$.

1. Compute $(\hat{x}_{n+}, \Pi_{n+})$ by EKF;
2. initialize an ICNN g_θ and define the ansatz

$$V_{n+|\theta}(x) = \frac{1}{2}\langle \nabla g_\theta(x) - \hat{x}_{n+}, \Pi_{n+}^{-1}(\nabla g_\theta(x) - \hat{x}_{n+})\rangle + \tilde{r}_{n+}; \quad (16)$$

3. for $k \in \{1, \cdots, K\}$, sample $X^{(k)} \sim \mathcal{N}(\hat{x}_{n+}, \Pi_{n+})$ and define $\mu_{n+}^{(K)} = \frac{1}{K}\sum_{k=1}^{K} \delta_{X^{(k)}}$;
4. define the least-squares loss function

$$L_{n+}(\theta) = \mathbb{E}_{X\sim\mu_{n+}^{(K)}}\left\{|V_{n+|\theta}(X) - V_{n+}(X)|^2\right\}; \quad (17)$$

5. train the NN: $\theta_{n+}^* \in \arg\min L_{n+}(\theta)$, set $V_{n+} := V_{n+|\theta_{n+}^*}$;

We rely on EKF and correct the nonlinearity with OT.

EKF-OT (correction step; prediciton step not shown but similar)

Let $n \in \mathbb{N}$, and assume we have computed approximations $(\hat{\mathrm{x}}_{n-}, \Pi_{n-})$.

1. Compute $(\hat{x}_{n+}, \Pi_{n+})$ by EKF;
2. initialize an ICNN g_θ and define the ansatz
$$V_{n+|\theta}(x) = \frac{1}{2}\langle \nabla g_\theta(x) - \hat{x}_{n+}, \Pi_{n+}^{-1}(\nabla g_\theta(x) - \hat{x}_{n+})\rangle + \tilde{r}_{n+}; \quad (16)$$
3. for $k \in \{1, \cdots, K\}$, sample $X^{(k)} \sim \mathcal{N}(\hat{x}_{n+}, \Pi_{n+})$ and define $\mu_{n+}^{(K)} = \frac{1}{K}\sum_{k=1}^{K} \delta_{X^{(k)}}$;
4. define the least-squares loss function
$$L_{n+}(\theta) = \mathbb{E}_{X\sim\mu_{n+}^{(K)}}\left\{|V_{n+|\theta}(X) - V_{n+}(X)|^2\right\}; \quad (17)$$
5. train the NN: $\theta^*_{n+} \in \arg\min L_{n+}(\theta)$, set $V_{n+} := V_{n+|\theta^*_{n+}}$;
6. compute $\hat{\mathrm{x}}_{n+} \in \arg\min V_{n+}(x)$ and $\Pi_{n+} = \nabla^2 V_{n+}(\hat{\mathrm{x}}_{n+})^{-1}$.

Outline

Linear tests

First linear test :

$$f = 0, \quad h(x) = x, \quad \nu = 0.$$

Let us set $(R^{-1}, S^{-1}, \Lambda_0, \tau, \hat{x}_0) = (8\,\mathrm{Id}, \mathrm{Id}, \mathrm{Id}, 0.1, 0.5)$, with constant target trajectory $y(t) = 2$.

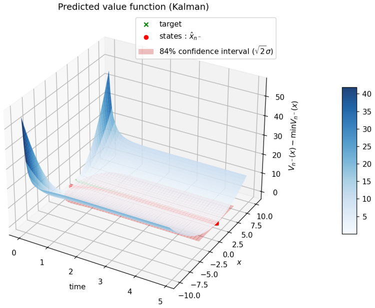

Figure: Kalman filter

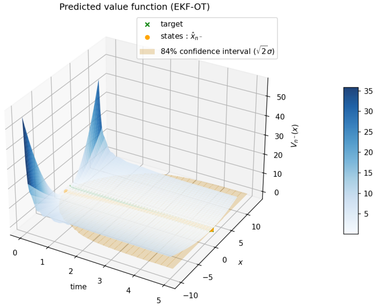

Figure: EKF-OT algorithm

Linear tests

First linear test :

$$f = 0, \quad h(x) = x, \quad \nu = 0.$$

Let us set $(R^{-1}, S^{-1}, \Lambda_0, \tau, \hat{x}_0) = (8\,\mathrm{Id}, \mathrm{Id}, \mathrm{Id}, 0.1, 0.5)$, with constant target trajectory $y(t) = 2$.

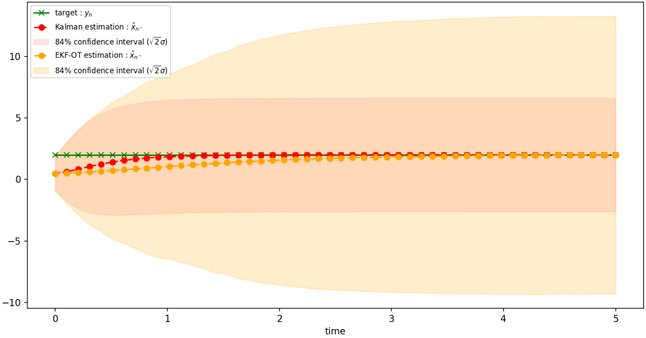

Figure: estimations

Linear tests

Second linear test :

$$f = 0, \quad h(x) = x, \quad \nu(t) = 0.6(\cos(4t) + \cos(8t)).$$

Let us set $(R^{-1}, S^{-1}, \Lambda_0, \tau, \hat{\mathrm{x}}_0) = (8\,\mathrm{Id}, \mathrm{Id}, \mathrm{Id}, 0.1, 0.5)$, with target trajectory $y(t) = 2 + \nu(t)$.

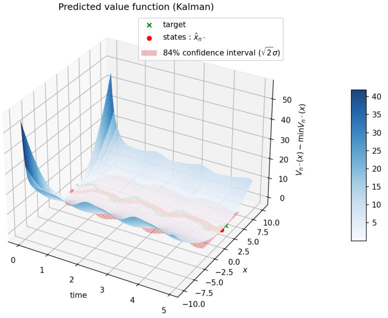

Figure: Kalman filter

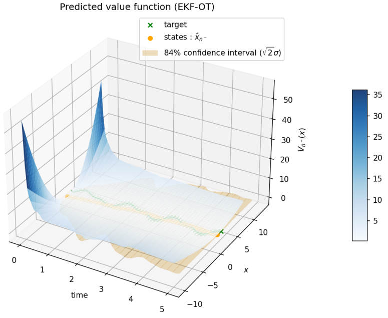

Figure: EKF-OT algorithm

Linear tests

Second linear test :

$$f = 0, \quad h(x) = x, \quad \nu(t) = 0.6(\cos(4t) + \cos(8t)).$$

Let us set $(R^{-1}, S^{-1}, \Lambda_0, \tau, \hat{x}_0) = (8\,\mathrm{Id}, \mathrm{Id}, \mathrm{Id}, 0.1, 0.5)$, with target trajectory $y(t) = 2 + \nu(t)$.

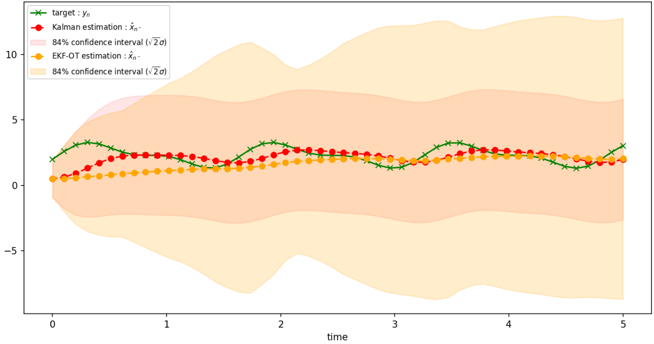

Figure: estimations

Nonlinear test

Nonlinear test (nonlinear observation function):

$$f = 0, \quad h(x) = x^2, \quad \nu = 0.$$

Let us set $(R^{-1}, S^{-1}, \Lambda_0, \tau, \hat{x}_0) = (\mathrm{Id}, \mathrm{Id}, \mathrm{Id}, 0.1, -0.5)$, with constant target trajectory $y(t) = 2$.

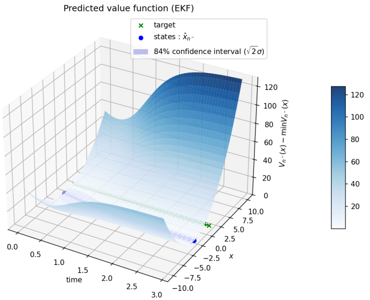

Figure: Extended Kalman Filter

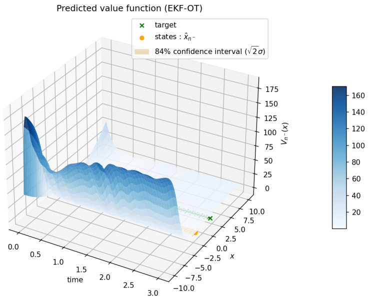

Figure: EKF-OT algorithm

Nonlinear tests

Nonlinear test :

$$f = 0, \quad h(x) = x^2, \quad \nu = 0.$$

Let us set $(R^{-1}, S^{-1}, \Lambda_0, \tau, \hat{x}_0) = (\mathrm{Id}, \mathrm{Id}, \mathrm{Id}, 0.1, -0.5)$, with target trajectory $y(t) = 2$.

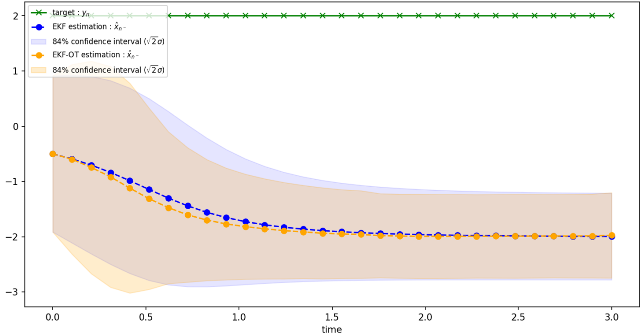

Figure: estimations

Conclusions and perspectives

- Gridless method, but there is a tradeoff (dimension/number of parameters) that is not satisfied in low dimensions;
- our method seems to agree on simple examples in 1D $\implies$ we have to go beyond to assert scalability;
- focus on examples where EKF **fails**, can the OT algorithm converge where EKF cannot ?
- Neural Networks are involved: how to fine tune them ?
- we lack a convergence result for now, large depth/width limit ?

- Amos, Brandon, Lei Xu, and J Zico Kolter (2017). "Input convex neural networks". In: *International conference on machine learning*. PMLR, pp. 146–155.
- Crandall, Michael G. and Pierre-Louis Lions (1983). "Viscosity solutions of Hamilton-Jacobi equations". In: *Trans. Amer. Math. Soc.* 277.1, pp. 1–42. ISSN: 0002-9947,1088-6850. DOI: 10.2307/1999343. URL: https://doi-org.extranet.enpc.fr/10.2307/1999343.
- Fleming, Wendell H (1997). "Deterministic nonlinear filtering". In: *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze* 25.3-4, pp. 435–454.
- Kalman, R. E. and R. S. Bucy (1961). "New results in linear filtering and prediction theory". In: *Trans. ASME Ser. D. J. Basic Engrg.* 83, pp. 95–108. ISSN: 0021-9223.
- Moireau, Philippe (2018). "A discrete-time optimal filtering approach for non-linear systems as a stable discretization of the Mortensen observer". In: *ESAIM: Control, Optimisation and Calculus of Variations* 24.4, pp. 1815–1847.
- Mortensen, R. E. (1968). "Maximum-likelihood recursive nonlinear filtering". In: *J. Optim. Theory Appl.* 2.6, pp. 386–394. ISSN: 0022-3239,1573-2878. DOI: 10.1007/BF00925744. URL: https://doi-org.extranet.enpc.fr/10.1007/BF00925744.