Nonlinear Filtering for Markovian Processes in Partially Known and Observed Orthogonal Subspaces.

Christophe Baehr and Antoine Campi

Météo-France / CNRS - CNRM / GAME UMR3589
Mathematical Institute of Toulouse III University

MANU-LEFE
Colloque National Assimilation de Données
Wednesday, December 03th 2014

Short summary

Classical nonlinear filter with particle approximations

Classical nonlinear filtering

- Let be $X_n \in \mathbb{L}^2$ with a dynamical equation

$$X_{n+1} = F^M_{n+1}(X_n) + W_{n+1}$$

where W_{n+1} is a Martingale process.

- The process X_n is partially observed by the process Y_n with an observation equation

$$Y_n = H(X_n) + V_n$$

where V_n is a noise process with a known pdf G_n

- The nonlinear filtering problem is the estimation of the conditional probability distribution $\hat{\eta}_n = \mathbb{E}[X_n | Y_{[0:n]}]$

- Using the Baye's decomposition, we get a sequential algorithm.

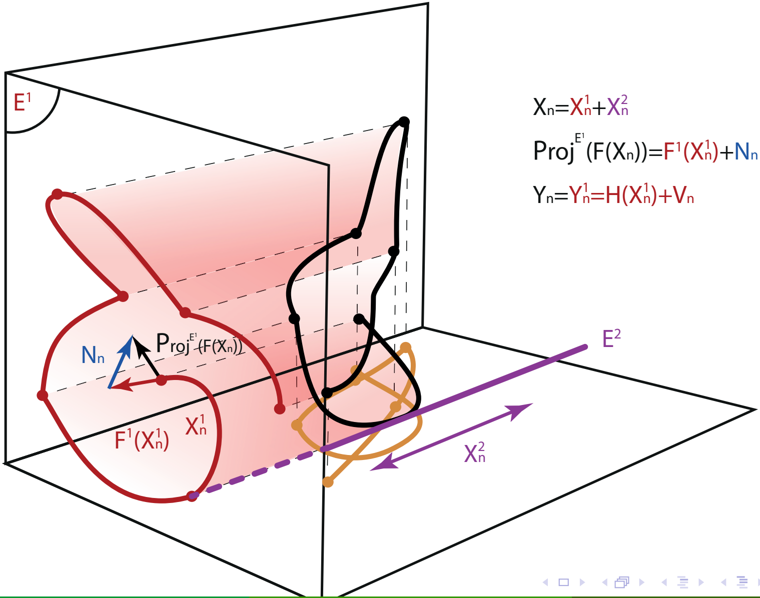
- For the general case, a particle approximation of the nonlinear filtering is used.

Classical nonlinear filtering

- Some definitions :
  - The Markovian evolution : $M_{n+1}(x, dy) = \mathbb{P}(X_{n+1} \in dy | X_n = x)$.
  - A potential function : G_{n+1} such that $\forall x \in \mathbb{C}(K, F),\ 0 \leq G_{n+1}(x) \leq 1$
  - The Bayes-Boltzmann-Gibbs transformation :
    $\Psi_{n+1}(\eta)(dx) = \dfrac{G_{n+1}(x)}{\eta(G_{n+1})}\eta(dx)$
  - Update selection : $S_{n+1,\eta_{n+1}}(x, dy)$ such that $\eta S_{n+1,\eta} = \Psi_{n+1}(\eta)$
    1. SIR : $S_{n+1,\eta}(x, dy) = \Psi_{n+1}(\eta)(dy)$
    2. Genetic Algorithm :
       $S_{n+1,\eta}(x, dy) = G_{n+1}(x)\delta_x + (1 - G_{n+1}(x))\ \Psi(\eta)(dy)$
  - The filtering Mc Kean kernel : $K_{n+1,\eta} = S_{n,\eta} M_{n+1}$

Filtering in Orthogonal Subspaces

Dynamics in Orthogonal Subspaces



$X_n = X_n^1 + X_n^2$

$\text{Proj}^{E^1}(F(X_n)) = F^1(X_n^1) + N_n$

$Y_n = Y_n^1 = H(X_n^1) + V_n$

Hypotheses

- Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a complete filtered probability space.
- Let be the polish space $E_n \subset \mathbb{L}^2 = E_n^1 \oplus E_n^2$ endowed with the σ-algebra $\mathcal{E}$ and a prehilbertian structure such that $E_n^1 \perp\!\!\!\perp E_n^2$
- Let be the random process $X_n = (X_n^1, X_n^2)$ where $X_n^1 \in E_n^1$ and $X_n^2 \in E_n^2$
- We denoted the probability law of the state X_n by $\eta_n = \eta_n^1 \otimes \eta_n^2$
- We assume that X_n have the dynamical model :

$$\begin{aligned} X_{n+1} &= F_{n+1}(X_n) + W_n \\ &= F_{n+1}(X_n^1, X_n^2) + W_n \\ &= (F_{n+1}^1(X_n^1, X_n^2)\ ,\ F_{n+1}^2(X_n^1, X_n^2)) + W_n \end{aligned}$$

- and we consider the following decomposition :

$$\begin{aligned} F_{n+1}^1(X_n^1, X_n^2) &= F_{n+1}^{1,1}(X_n^1) + F_{n+1,X_n^1}^{2,1}(X_n^2) \\ F_{n+1}^2(X_n^1, X_n^2) &= F_{n+1}^{1,2}(X_n^1) + F_{n+1,X_n^1}^{2,2}(X_n^2) \end{aligned}$$

➔ The Markovian transition kernel of the state X_n is

$$\begin{aligned} M_{n+1}(x, dz) &= M_{n+1}((x^1, x^2), d(z^1, z^2)) \\ &= \mathbb{P}(X_{n+1}^1 \in dz^1 | X_n = x) \otimes \mathbb{P}(X_{n+1}^2 \in dz^2 | X_n = x) \end{aligned}$$

The selection occcurs only on E_n^1

- Let be the observation process $Y_n = Y_n^1 = H_n(Proj^{E_n^1} X_n) + V_n = H_n^1(X_n) + V_n$

Proposition

The selection kernel for blind orthogonal subspaces is

$$S_{n,\eta_n}(x, dz) = S^1_{n,\eta_n^1}(x, dz^1) \otimes \eta_n^2(dz^2)$$

Proof :
There is no observation on E_n^2, then the potential G_n of the state X_n which is defined by the likelihood of y^1 given x^1 may be written $G_n(x) = G_n^1(x_n^1)$ and

$$\eta_n(G_n^1) = \int_x G_n^1(x^1)\eta_n(dx) = \int_{x^1} G_n^1(x_n^1)\eta_n^1(dx^1)$$

Then the selection kernel S_{n,η_n} is

$$S_{n,\eta_n}(x, dz) = \frac{G_n(z)}{\eta_n(G_n)}\eta_n(dz) = \frac{G_n^1(z)}{\eta_n^1(G_n^1)}\eta_n^1(dz^1)\otimes\eta_n^2(dz^2) = S^1_{n,\eta_n^1}(x, dz^1)\otimes\eta_n^2(dz^2) \quad \blacksquare$$

The total Markovian transition

Theorem

Using the general hypotheses and the previous definitions, it yields

$$M_{n+1}(x, dy) = M^1_{n+1}(x, dy^1) M^2_{n+1}(x, dy^2) = \int_\xi M^{1,\cdot}_{n+1}(x^1, d\xi) M^{2,\cdot}_{n+1, F^{2,\cdot}_{n+1,x^1}(x^2)}(\xi, dy)$$

- We recall the general dynamics

$$X_{n+1} = \left(F^1_{n+1}(X^1_n, X^2_n)\ ,\ F^2_{n+1}(X^1_n, X^2_n)\right) + W_n$$

where we get :

$$\begin{aligned} F^1_{n+1}(X^1_n, X^2_n) &= F^{1,1}_{n+1}(X^1_n) + F^{2,1}_{n+1,X^1_n}(X^2_n) \\ F^2_{n+1}(X^1_n, X^2_n) &= F^{1,2}_{n+1}(X^1_n) + F^{2,2}_{n+1,X^1_n}(X^2_n) \end{aligned}$$

- The effects of the orthogonal subspaces express themselves through a parametered transition kernel.

Some necessary definitions

First we defined the different kernels

$$M^{1,\cdot}_{n+1}(x^1, d\xi) = M^{1,1}_{n+1}(x^1, d\xi^1) M^{1,2}_{n+1}(x^1, d\xi^2)$$

$$M^{2,\cdot}_{n+1, F^{2,\cdot}_{n+1,x^1}(x^2)}(\xi, dy) = M^{2,1}_{n+1, F^{2,1}_{n+1,x^1}(x^2)}(\xi^1, dy^1) M^{2,2}_{n+1, F^{2,2}_{n+1,x^1}(x^2)}(\xi^2, dy^2)$$

and

$$M^{\cdot,1}_{n+1}(x, dy^1) = \int_{\xi^1} M^{1,1}_{n+1}(x^1, d\xi^1) M^{2,1}_{n+1, F^{2,1}_{n+1,x^1}(x^2)}(\xi^1, dy^1)$$

$$M^{\cdot,2}_{n+1}(x, dy^2) = \int_{\xi^2} M^{1,2}_{n+1}(x^1, d\xi^2) M^{2,2}_{n+1, F^{2,2}_{n+1,x^1}(x^2)}(\xi^2, dy^2)$$

with the initial definitions

$$M^{1,1}_{n+1}(x^1, d\xi^1) = \mathbb{P}\left(F^{1,1}_{n+1}(X^1_n) \in d\xi^1 | X^1_n = x^1 \right)$$

$$M^{2,1}_{n+1, F^{2,1}_{n+1,x^1}(x^2)}(\xi^1, dy^1) = \mathbb{P}\left(X^1_{n+1} \in dy^1 | X_n = x, F^{1,1}_{n+1}(X^1_n) = \xi^1 \right)$$

$$M^{1,2}_{n+1}(x^1, d\xi^2) = \mathbb{P}\left(F^{1,2}_{n+1}(X^1_n) \in d\xi^2 | X^1_n = x^1 \right)$$

$$M^{2,2}_{n+1, F^{2,2}_{n+1,x^1}(x^2)}(\xi^2, dy^2) = \mathbb{P}\left(X^2_{n+1} \in dy^2 | X_n = x, F^{1,2}_{n+1}(X^1_n) = \xi^2 \right)$$

The total Markovian transition

Theorem

Using the general hypotheses and the previous definitions, it yields

$$M_{n+1}(x, dy) = M^1_{n+1}(x, dy^1) M^2_{n+1}(x, dy^2) = \int_\xi M^{1,\cdot}_{n+1}(x^1, d\xi) M^{2,\cdot}_{n+1, F^{2,\cdot}_{n+1,x^1}(x^2)}(\xi, dy)$$

Proof :

The result is a direct calculation using the previous light definitions and the definition of M_{n+1} :

$$M_{n+1}(x, dz) = \mathbb{P}(X^1_{n+1} \in dz^1 | X_n = x) \otimes \mathbb{P}(X^2_{n+1} \in dz^2 | X_n = x)$$

■

The filtering algorithm is determined by the McKean kernel

Theorem

The filtering McKean kernel $K_{n+1,\eta_n}(x,dz) = \int_\xi S_{n,\eta_n}(x,d\xi)M_{n+1}(\xi,dz)$ *according to the orthogonal subspaces is given by*

$$\begin{aligned} K_{n+1,\eta_n}(x,dy) &= S_{n,\eta_n^1} M^{1,1}_{n+1} \mathbb{M}^{2,1}_{n+1,\eta^2_n,.^1}(x,dy^1) \otimes S_{n,\eta_n^1} M^{1,2}_{n+1} \mathbb{M}^{2,2}_{n+1,\eta^2_n,.^1}(x,dy^2) \\ &= K^1_{n+1,\eta_n}(x,dy^1) \otimes K^2_{n+1,\eta_n}(x,dy^2) \end{aligned}$$

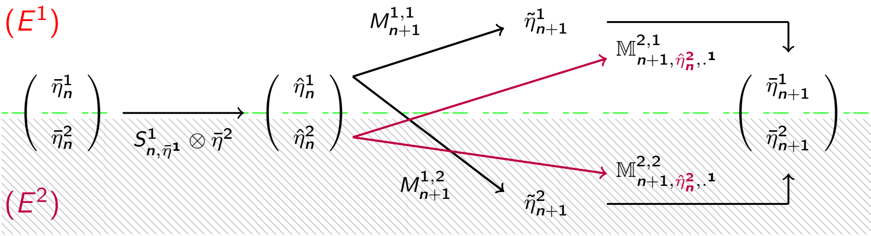
where we denote

$$\mathbb{M}^{2,i}_{n+1,\eta_n^2} = \eta_n^2 M^{2,i}_{n+1,F^{2,i}_{n+1}(\bullet)}$$

Proof :
The proof is only the combination of the previous results ■

- $\mathbb{M}^{2,i}_{n+1,\eta_n^2}$ is called the novation®. The novation® is the feedback of the orthogonal subspaces between themselves.
- **Note :** If the two subspaces are observed, the McKean evolution have a shape like a Rao-Blawellized particle filter.

The filtering algorithm considering the 2 susbpaces



where $\left| \begin{array}{l} \bar{\eta}^1_n \\ \bar{\eta}^2_n \end{array} \right.$ are the predictor measures, $\left| \begin{array}{l} \hat{\eta}^1_n \\ \hat{\eta}^2_n \end{array} \right.$ are the filtered distributions and $\left| \begin{array}{l} \tilde{\eta}^1_n \\ \tilde{\eta}^2_n \end{array} \right.$ the prediction before the novation® corrections.

Questionning about the Novation® kernel

- How to get an estimation of the Novation® kernel $\mathbb{M}^{2,1}_{n+1,\eta^2_n}$?
- How to get an interpretation for the total prediction step $M^{1,1}_{n+1}\mathbb{M}^{2,1}_{n+1,\eta^2_n}$?
- How to control the Novation® estimation error ?

About the Novation® estimation

Giving some material for the Novation® estimation

➔ In order to suggest an estimation algorithm for the Novation®, we begin with some remarks. First we assume that we have an estimation algorithm for the Novation®.

- The Novation® process depends on η_n^2 which is unreachable.
- Due to the dynamical structure, the Novation® estimation algorithm have no other choice than extracts information from the observation.
- Since the Novation® is a process, the information is, at the first order, a conditional average given the observation.

Giving some material for the Novation® estimation

- Have an estimation algorithm for the Novation® means that we have a random process $N_n(X_n^1, Y_n)$ such that for any $\delta > 0$

$$\|N_{n+1}(X_{n+1}^1, Y_{n+1}) - F_{n+1,X_n^1}^{2,1}(X_n^2)\| \leq \delta$$

  and there is a transition kernel $\mathcal{C}_{n+1}^{2,1} \sim \mathbb{M}_{n+1,\eta_n^2}^{2,1}$

- We assume that the Novation® is p-integrable and it exists a p-integrable measure μ such that

$$\|M_{n+1}^{1,1}\mathcal{C}_{n+1}^{2,1} - M_{n+1}^{1,1}\mathbb{M}_{n+1,\eta_n^2}^{2,1}\| \leq \|\mu\|$$

  $\| \bullet \|$ is the Total Variation norm.

- Using this little assumptions, we have proved that we may control the L^p errors of the Novation® estimation algorithm.

Estimation of the Novation® in a particular case

Application to the Gaussian case

- We assume that the Novation® is Gaussian.
- It means that we have a spatial average to determine and a covariance.
- This is the frame of the Particle Filter with imperfect model.
- We suggest to learn the conditional average by a variational minimization.
- The error covariance matrices used in the minimization are learned .... somewhere else ...

Application to the Gaussian case

Proposition

We assume that the observation function H is h-Lipschitz, the dynamical noise W_n follows a $\mathcal{N}(0, \sigma_W^2)$ and the observational noise V_n is $\mathcal{N}(0, \sigma_V^2)$ distributed. We denote P_n the error covariance matrix. Then, for any measure ν, it yields :

$$\|\nu M_n^{1,1}\mathcal{C}_n^{2,1} - \nu M_n^{1,1}\mathbb{M}_{n+1,\eta_n^2}^{2,1}\| \leq \|\mathcal{N}(\delta, (1+|1-h^2 P_n|\sigma_w^2 + P_n\sigma_v^2)\|$$

Application to the Gaussian case

- How to learn the P_n error covariance matrix ?
  - For low dimension systems ($<$ some thousands) $\longrightarrow$ use the well known Island Particle Method.
  - For a Gaussian and non-linear world $\longrightarrow$ use the well known Variational Interacting Filter (interaction through an importance resampling) .
  - For a Gaussian and non-linear world in very high dimension $\longrightarrow$ use the well known Unscented/Ensemble Variational Filter.
  - Other methods ....

Application to the Gaussian case

- For this particular case the non linear filter is approximated by a particle filter including a variational minimization which learn the blind subspace feedback average.

- We consider the particle system $(\hat{X}^i_{n-1})_{i=1}^N$, initialized, conveyed and filtered since the step $n-1$.

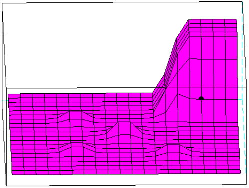
- The variational minimization is performed on the mean of the predicted particles $\tilde{X}^i_n$ and is used to get the corrected prediction particle set $\overline{X}^i_n$.

- Then, for the time step n, the algorithm is :

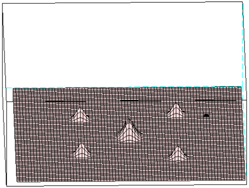
$$\hat{X}^i_{n-1} \xrightarrow{Prediction} \tilde{X}^i_n \xrightarrow{Mean} \tilde{X}_n \xrightarrow{VarMini} \overline{X}_n \xrightarrow{Debiasing} \overline{X}^i_n \xrightarrow{Selection} \hat{X}^i_n$$

A first numerical application of the Novation® estimation

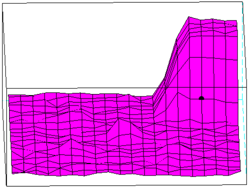
An application based on the Shallow Water Equation



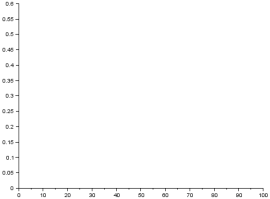
Total Process



Process in Orthogonal Subspace



Filtered Process



Height of the floating point

Dimensions : 1200   Nb of Particules : 50

A short remark about the dimension

- The average learning step reduces harshly the number of necessary particles.
- Indeed the efforts lay in the learning of the error covariance matrices.
- Numerical example : 2 Layers Quasi-Geostrophic Model, 10 particles, dimension 3000.
- Why so few particles ?

A short remark about the dimension

➔ We suggest a work on the structure functions :

Proposition

For two points a and b we compute the covariance

$$\begin{aligned} S_n^{1,a,b} &= \mathbb{E}(X_n^{1,a} X_n^{1,b} \mid Y_n) \\ &= \tilde{S}_n^{1,a,b} \quad E^1 \textit{ model structure function} \\ &+ S_n^{2\to 1,a,b} \quad \textit{Novation structure function} \\ &+ 2.S_n^{2x1,a,b} \quad \textit{Covariances novation-model} \end{aligned}$$

Then, it exists a constant C^2 such that $\|\overline{S}_n^{1,a,b} - S_n^{1,a,b}\| \le C^2$ where $\overline{S}_n^{1,a,b}$ is the estimation of $S_n^{1,a,b}$ using the Novation estimation algorithm.

Moreover if the Novation estimation error goes to zero, C^2 goes also to zero.

A short remark about the dimension

- Then there is no more question of dimension ... or almost no more.
- Indeed for a set of d points $X_n^{1\cdots d}$ we may write

$$\begin{aligned}\mathbb{P}(\tilde{X}_n^{1..d} \in dx^{1..d}|Y_n^{1..d}) &= \int_z \mathbb{P}(\tilde{X}_n^{1..d} \in dx^{1..d}|Y_n^{1..d}, \tilde{X}_n^1 = z)\mathbb{P}(\tilde{X}_n^1 \in dz|Y_n^{1..d}) \\ &= \mathbb{P}(\tilde{X}_n^{2..d} \in dx^{2..d}|Y_n^{1..d}, \tilde{X}_n^1 = x^1)\mathbb{P}(\tilde{X}^1 \in dx^1|Y_n^1)\end{aligned}$$

And if the structure functions are exactly determined, it yields

$$\mathbb{P}(\tilde{X}_n^{1..d} \in dx^{1..d}|Y_n^{1..d}) = \bigotimes_{i=2}^{d} \delta_{\tilde{x}^i} \ . \ \mathbb{P}(\tilde{X}^1 \in dx^1|Y^1)$$

- Then a reduce set of particles is necessaty to learn $\mathbb{P}(\tilde{X}^1 \in dx^1|Y^1)$.
- In nominal conditions, the structure functions are not entirely determined and a bigger set of particles is necessary...

Thanks for your attention

Bayesianity

➔ We assume that the total state is x_n and we denote $z_n = -N_n$. Then $x_n + z_n$ is the prediction in E^1 without Novation correction.

$$
\begin{aligned}
& p(x_n \mid y_{[0,n]}) \\
= & \frac{p(y_n \mid x_n, y_{[0,n-1]}) p(x_n \mid y_{[0,n-1]})}{p(y_n \mid y_{[0,n-1]})} \\
= & \frac{p(y_n \mid x_n, y_{[0,n-1]})}{p(y_n \mid y_{[0,n-1]})} \int p(x_n \mid x_n + z_n, y_{[0,n-1]}) p(x_n + z_n \mid y_{[0,n-1]}) \\
= & \frac{p(y_n \mid x_n, y_{[0,n-1]})}{p(y_n \mid y_{[0,n-1]})} \int p(x_n \mid x_n + z_n, y_{[0,n-1]}) \\
& \left(\int p(x_n + z_n \mid x_{n-1}, y_{[0,n-1]}) p(x_{n-1} \mid y_{[0,n-1]}) \right)
\end{aligned}
$$

and

- $\tilde{\eta}_n = p(x_n | y_{[0:n-1]})$, $\overline{\eta}_n = p(x_n + z_n | y_{[0,n-1]})$ et $\hat{\eta}_n = p(x_n | y_{[0,n]})$,
- $M_n = p(x_n \mid x_{n-1}, y_{[0,n-1]})$,
- $M_n = M_n^{1,1} \mathcal{C}_n^{2,1}$ avec $M_n^{1,1} = p(x_n + z_n \mid x_{n-1}, y_{[0,n-1]})$ et $\mathcal{C}_n^{2,1} = p(x_n \mid x_n + z_n, y_{[0,n-1]})$.