Adaptive Gaussian nested filter for joint parameter and state estimation in state-space models

Sara Pérez Vieites

CERI Systèmes Numériques,
IMT Nord Europe (Lille, France).

Joint work with
Joaquín Míguez (Universidad Carlos III de Madrid, Spain) and
Víctor Elvira (University of Edinburgh).

MCM 2023

Index

State-space model

We are interested in systems can be represented by **Markov state-space dynamical models**:

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{v}_t, \\ \boldsymbol{y}_t &= \boldsymbol{g}(\boldsymbol{x}_t, \boldsymbol{\theta}) + \boldsymbol{r}_t, \end{aligned}$$

- $\boldsymbol{f}$, $\boldsymbol{g}$: state transition function and observation function
- $\boldsymbol{v}_t$, $\boldsymbol{r}_t$: state and observation noises

In terms of a set of **relevant probability density functions (pdfs)**:

- Prior pdfs: $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ and $\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0)$
- Transition pdf of the state: $\boldsymbol{x}_t \sim p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{\theta})$
- Conditional pdf of the observation: $\boldsymbol{y}_t \sim p(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{\theta})$

State-space model

We are interested in systems can be represented by **Markov state-space dynamical models**:

$$
\begin{aligned}
\boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{v}_t, \\
\boldsymbol{y}_t &= \boldsymbol{g}(\boldsymbol{x}_t, \boldsymbol{\theta}) + \boldsymbol{r}_t,
\end{aligned}
$$

- $\boldsymbol{f}$, $\boldsymbol{g}$: state transition function and observation function
- $\boldsymbol{v}_t$, $\boldsymbol{r}_t$: state and observation noises

In terms of a set of **relevant probability density functions (pdfs)**:

- Prior pdfs: $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ and $\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0)$
- Transition pdf of the state: $\boldsymbol{x}_t \sim p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{\theta})$
- Conditional pdf of the observation: $\boldsymbol{y}_t \sim p(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{\theta})$

State estimation

Classical filtering methods:

Bayesian estimation of the state variables, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star)$, assuming $\boldsymbol{\theta}^\star$ is known.

Every time step t:

1. Predictive distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \theta^\star) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \theta^\star)p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}, \theta^\star)d\boldsymbol{x}_t \quad (1)$$

2. Likelihood: $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \theta^\star)$
3. Posterior/filtering distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \theta^\star) \propto (\boldsymbol{y}_t|\boldsymbol{x}_t, \theta^\star)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \theta^\star) \quad (2)$$

In practice, $\theta^\star$ is not known. It is needed to **estimate both θ and $\boldsymbol{x}_t$**, i.e., $p(\boldsymbol{x}_t, \theta|\boldsymbol{y}_{1:t})$.

State estimation

Classical filtering methods:

Bayesian estimation of the state variables, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star)$, assuming $\boldsymbol{\theta}^\star$ is known.

Every time step t:

1. Predictive distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}^\star) p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) d\boldsymbol{x}_t \quad (1)$$

2. Likelihood: $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}^\star)$
3. Posterior/filtering distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star) \propto (\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}^\star) p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) \quad (2)$$

In practice, $\boldsymbol{\theta}^\star$ is not known. It is needed to **estimate both $\boldsymbol{\theta}$ and $\boldsymbol{x}_t$**, i.e., $p(\boldsymbol{x}_t, \boldsymbol{\theta}|\boldsymbol{y}_{1:t})$.

State estimation

Classical filtering methods:

Bayesian estimation of the state variables, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star)$, assuming $\boldsymbol{\theta}^\star$ is known.

Every time step t:

1. Predictive distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}^\star) p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) d\boldsymbol{x}_t \quad (1)$$

2. Likelihood: $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}^\star)$
3. Posterior/filtering distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star) \propto (\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}^\star) p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) \quad (2)$$

In practice, $\theta^\star$ is not known. It is needed to **estimate both** θ **and** x_t, i.e., $p(x_t, \theta|y_{1:t})$.

State estimation

Classical filtering methods:

Bayesian estimation of the state variables, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star)$, assuming $\boldsymbol{\theta}^\star$ is known.

Every time step t:

1. Predictive distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}^\star) p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) d\boldsymbol{x}_t \tag{1}$$

2. Likelihood: $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}^\star)$
3. Posterior/filtering distribution:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta}^\star) \propto (\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}^\star) p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^\star) \tag{2}$$

In practice, $\boldsymbol{\theta}^\star$ is not known. It is needed to **estimate both θ and $\boldsymbol{x}_t$**, i.e., $p(\boldsymbol{x}_t, \theta|\boldsymbol{y}_{1:t})$.

State-of-the-art methods

Methods for Bayesian inference of both $\boldsymbol{\theta}$ and $\boldsymbol{x}_t$:

- **particle Markov chain Monte Carlo (PMCMC)**[1]
- **sequential Monte Carlo square ($\mathbf{SMC^2}$)**[2]

⟶ They can quantify the uncertainty or estimation error,
⟶ they can be applied to a broad class of models,
⟶ they provide theoretical guarantees,
⟶ they are batch techniques.
⟶ The computational cost becomes prohibitive in high-dimensional problems.

Motivation: Reduction of the computational complexity.

[1] Andrieu, Doucet, and Holenstein 2010.
[2] Chopin, Jacob, and Papaspiliopoulos 2011.

State-of-the-art methods

Methods for Bayesian inference of both $\boldsymbol{\theta}$ and $\boldsymbol{x}_t$:

- **particle Markov chain Monte Carlo (PMCMC)**[1]
- **sequential Monte Carlo square (SMC^2)**[2]

$\longrightarrow$ They can quantify the uncertainty or estimation error,
$\longrightarrow$ they can be applied to a broad class of models,
$\longrightarrow$ they provide theoretical guarantees,
$\longrightarrow$ they are batch techniques.
$\longrightarrow$ The computational cost becomes prohibitive in high-dimensional problems.

Motivation: Reduction of the computational complexity.

[1] Andrieu, Doucet, and Holenstein 2010.
[2] Chopin, Jacob, and Papaspiliopoulos 2011.

State-of-the-art methods

Methods for Bayesian inference of both $\boldsymbol{\theta}$ and $\boldsymbol{x}_t$:

- **particle Markov chain Monte Carlo (PMCMC)**[1]
- **sequential Monte Carlo square (SMC^2)**[2]

$\longrightarrow$ They can quantify the uncertainty or estimation error,
$\longrightarrow$ they can be applied to a broad class of models,
$\longrightarrow$ they provide theoretical guarantees,
$\longrightarrow$ they are batch techniques.
$\longrightarrow$ The computational cost becomes prohibitive in high-dimensional problems.

Motivation: Reduction of the computational complexity.

[1] Andrieu, Doucet, and Holenstein 2010.
[2] Chopin, Jacob, and Papaspiliopoulos 2011.

State-of-the-art methods

Methods for Bayesian inference of both $\boldsymbol{\theta}$ and $\boldsymbol{x}_t$:

- **particle Markov chain Monte Carlo (PMCMC)**[1]
- **sequential Monte Carlo square (SMC^2)**[2]

$\longrightarrow$ They can quantify the uncertainty or estimation error,
$\longrightarrow$ they can be applied to a broad class of models,
$\longrightarrow$ they provide theoretical guarantees,
$\longrightarrow$ they are batch techniques.
$\longrightarrow$ The computational cost becomes prohibitive in high-dimensional problems.

Motivation: Reduction of the computational complexity.

[1] Andrieu, Doucet, and Holenstein 2010.
[2] Chopin, Jacob, and Papaspiliopoulos 2011.

Index

Model inference

We aim at computing the joint posterior pdf $p(\boldsymbol{\theta}, \boldsymbol{x}_t | \boldsymbol{y}_{1:t})$, that can be written as

$$p(\boldsymbol{x}_t, \boldsymbol{\theta} | \boldsymbol{y}_{1:t}) = \underbrace{p(\boldsymbol{x}_t | \boldsymbol{\theta}, \boldsymbol{y}_{1:t})}_{2^{nd}\ \text{layer}} \underbrace{p(\boldsymbol{\theta} | \boldsymbol{y}_{1:t})}_{1^{st}\ \text{layer}}$$

$\longrightarrow$ The **key difficulty** in this class of models is the Bayesian estimation of the parameter vector $\boldsymbol{\theta}$.

Model inference

At every time step t:

1^{st} layer

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of }\boldsymbol{\theta}}$$

Filtering (given θ)

$$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of }\boldsymbol{x}}$$

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t,\boldsymbol{\theta})$$

$$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t})}_{\text{Post. pdf of }\boldsymbol{x}}$$

2^{nd} layer

$$p(\boldsymbol{y}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})$$

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})}_{\text{Post. pdf of }\boldsymbol{\theta}} \propto p(\boldsymbol{y}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})$$

Model inference

At every time step t:

1^{st} layer

$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{\theta}}$

Filtering (given θ)

2^{nd} layer

$\underbrace{p(\boldsymbol{x}_t|\theta, \boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{x}}$

$p(\boldsymbol{y}_t|\boldsymbol{x}_t, \theta)$

$\underbrace{p(\boldsymbol{x}_t|\theta, \boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{x}}$

$$p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta})p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1})d\boldsymbol{x}_t$$

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{\theta}} \propto p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})$$

Model inference

At every time step t:

1st layer

$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{\theta}}$

Filtering (given $\boldsymbol{\theta}$)

$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{x}}$

$p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta})$

$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{x}}$

2nd layer

$$p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}) p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) d\boldsymbol{x}_t$$

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{\theta}} \propto p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})$$

Model inference

At every time step t:

1st layer

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{\theta}}$$

Filtering (given $\boldsymbol{\theta}$)

$$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{x}} = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{\theta})p(\boldsymbol{x}_{t-1}|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})d\boldsymbol{x}_{t-1}$$

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t,\boldsymbol{\theta})$$

$$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{x}} \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t,\boldsymbol{\theta})p(\boldsymbol{x}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})$$

2nd layer

$$p(\boldsymbol{y}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t,\boldsymbol{\theta})p(\boldsymbol{x}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})d\boldsymbol{x}_t$$

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{\theta}} \propto p(\boldsymbol{y}_t|\boldsymbol{\theta},\boldsymbol{y}_{1:t-1})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})$$

Model inference

At every time step t:

1^{st} layer

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{\theta}}$$

Filtering (given $\boldsymbol{\theta}$)

$$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1})}_{\text{Pred. pdf of } \boldsymbol{x}} = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) p(\boldsymbol{x}_{t-1}|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) d\boldsymbol{x}_{t-1}$$

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta})$$

$$\underbrace{p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{x}} \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}) p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1})$$

2^{nd} layer

$$p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}) p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) d\boldsymbol{x}_t$$

$$\underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})}_{\text{Post. pdf of } \boldsymbol{\theta}} \propto p(\boldsymbol{y}_t|\boldsymbol{\theta}, \boldsymbol{y}_{1:t-1}) p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1})$$

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC (N_θ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_x$:

SMC (N_x samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_x$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_x} \tilde{u}_t^{i,j}}$

- Likelihood of $\boldsymbol{\theta}^i$: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\boldsymbol{\theta}_t^i}(d\boldsymbol{\theta})$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\theta^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_x$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_x} \tilde{u}_t^{i,j}}$

- Likelihood of $\boldsymbol{\theta}^i$: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\theta|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\boldsymbol{\theta}_t^i}(d\boldsymbol{\theta})$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_x$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_x} \tilde{u}_t^{i,j}}$

- Likelihood of $\boldsymbol{\theta}^i$: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\theta|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\theta_t^i}(d\boldsymbol{\theta})$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_x$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_x} \tilde{u}_t^{i,j}}$

- Likelihood of θ^i: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\theta|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\theta_t^i}(d\theta)$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}}$

- Likelihood of θ^i: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\theta|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\theta_t^i}(d\theta)$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}}$

- Likelihood of $\boldsymbol{\theta}^i$: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\theta|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\theta_t^i}(d\theta)$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}}$

- Likelihood of $\boldsymbol{\theta}^i$: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_\theta} w_t^i \delta_{\boldsymbol{\theta}_t^i}(d\boldsymbol{\theta})$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_\theta} \tilde{w}_t^i}$.

Naive importance sampling approximation

Initialisation: Draw $\{\boldsymbol{\theta}^i\}_{i=1}^{N_\theta}$ from $p(\boldsymbol{\theta})$

At $t \geq 1$ and for every $\boldsymbol{\theta}^i$, $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $j = 1, \ldots, N_{\boldsymbol{x}}$:

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}^i)$

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \boldsymbol{\theta}^i)$
- Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}}$

- Likelihood of $\boldsymbol{\theta}^i$: $\tilde{w}_t^i = w_{t-1}^i \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \tilde{u}_t^{i,j} \right)$
- Then, $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) = \sum_{i=1}^{N_{\boldsymbol{\theta}}} w_t^i \delta_{\boldsymbol{\theta}_t^i}(d\boldsymbol{\theta})$, with $w_t^i = \frac{\tilde{w}_t^i}{\sum_{i=1}^{N_{\boldsymbol{\theta}}} \tilde{w}_t^i}$.

After several time steps the filter degenerates.

Nested particle filter (NPF)[3]

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

- Jittering: Draw $\bar{\boldsymbol{\theta}}_t^i \sim \kappa_{N_{\boldsymbol{\theta}}}(\boldsymbol{\theta}|d\boldsymbol{\theta}_{t-1}^i)$

SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \bar{\boldsymbol{\theta}}_t^i)$

Given $\bar{\boldsymbol{\theta}}_t^i$, for $j = 1, \ldots, N_{\boldsymbol{x}}$:

- Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\bar{\boldsymbol{\theta}}_t^i, \boldsymbol{y}_{1:t-1})$
- Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \bar{\boldsymbol{\theta}}_t^i)$
- Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}}$

- Likelihood of $\bar{\boldsymbol{\theta}}_t^i$: $\tilde{w}_t^i = \frac{1}{N_{\boldsymbol{x}}} \sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}$
- Resampling: for $l = 1, \ldots, N_{\boldsymbol{\theta}}$, $\{\boldsymbol{\theta}_t^i, \{\boldsymbol{x}_t^{i,j}\}_{1 \le j \le N_{\boldsymbol{x}}}\} = \{\bar{\boldsymbol{\theta}}_t^l, \{\tilde{\boldsymbol{x}}_t^{l,j}\}_{1 \le j \le N_{\boldsymbol{x}}}\}$ with prob. w_t^l, so that $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{i=1}^{N_{\boldsymbol{\theta}}} \delta_{\boldsymbol{\theta}_t^i}(d\boldsymbol{\theta})$

[3] Crisan and Miguez 2017.

Nested particle filter (NPF)[3]

SMC ($N_{\boldsymbol{\theta}}$ samples) to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$

For $i = 1, \ldots, N_{\boldsymbol{\theta}}$:

- Jittering: Draw $\bar{\boldsymbol{\theta}}_t^i \sim \kappa_{N_{\boldsymbol{\theta}}}(\boldsymbol{\theta}|d\boldsymbol{\theta}_{t-1}^i)$

  SMC ($N_{\boldsymbol{x}}$ samples) to approximate $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \bar{\boldsymbol{\theta}}_t^i)$

  Given $\bar{\boldsymbol{\theta}}_t^i$, for $j = 1, \ldots, N_{\boldsymbol{x}}$:

  - Draw $\bar{\boldsymbol{x}}_t^{i,j} \sim p(\boldsymbol{x}_t|\bar{\boldsymbol{\theta}}_t^i, \boldsymbol{y}_{1:t-1})$
  - Weights: $\tilde{u}_t^{i,j} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{i,j}, \bar{\boldsymbol{\theta}}_t^i)$
  - Resampling: for $m = 1, \ldots, N_{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}_t^{i,j} = \bar{\boldsymbol{x}}_t^{i,m}$ with prob. $u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}}$

- Likelihood of $\bar{\boldsymbol{\theta}}_t^i$: $\tilde{w}_t^i = \frac{1}{N_{\boldsymbol{x}}} \sum_{j=1}^{N_{\boldsymbol{x}}} \tilde{u}_t^{i,j}$
- Resampling: for $l = 1, \ldots, N_{\boldsymbol{\theta}}$, $\{\boldsymbol{\theta}_t^i, \{\boldsymbol{x}_t^{i,j}\}_{1 \le j \le N_{\boldsymbol{x}}}\} = \{\bar{\boldsymbol{\theta}}_t^l, \{\tilde{\boldsymbol{x}}_t^{l,j}\}_{1 \le j \le N_{\boldsymbol{x}}}\}$ with prob. w_t^l, so that $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{i=1}^{N_{\boldsymbol{\theta}}} \delta_{\boldsymbol{\theta}_t^i}(d\boldsymbol{\theta})$

High computational complexity: $N_{\boldsymbol{\theta}} \times N_{\boldsymbol{x}}$ samples.

[3] Crisan and Miguez 2017.

Family of nested filters

1. **Nested particle filters (NPFs)**[4].
   - Both layers ⟶ Sequential Monte Carlo (SMC) methods
2. **Nested hybrid filters (NHFs)**[5].
   - θ-layer ⟶ Monte Carlo-based methods (e.g., SMC or SQMC)
   - x-layer ⟶ Gaussian techniques (e.g., EKFs or EnKFs)
3. **Nested Gaussian filters (NGFs)**[6].
   - θ-layer ⟶ Deterministic sampling methods (e.g., UKF).
   - x-layer ⟶ Gaussian techniques (e.g., EKFs or EnKFs).

[4] Crisan and Míguez 2018.
[5] Pérez-Vieites, Mariño, and Míguez 2017.
[6] Pérez-Vieites and Míguez 2021.

Family of nested filters

1. **Nested particle filters (NPFs)**[4].
   - Both layers $\longrightarrow$ Sequential Monte Carlo (SMC) methods
2. **Nested hybrid filters (NHFs)**[5].
   - $\boldsymbol{\theta}$-layer $\longrightarrow$ Monte Carlo-based methods (e.g., SMC or SQMC)
   - $\boldsymbol{x}$-layer $\longrightarrow$ Gaussian techniques (e.g., EKFs or EnKFs)
3. **Nested Gaussian filters (NGFs)**[6].
   - $\boldsymbol{\theta}$-layer $\longrightarrow$ Deterministic sampling methods (e.g., UKF).
   - $\boldsymbol{x}$-layer $\longrightarrow$ Gaussian techniques (e.g., EKFs or EnKFs).

[4] Crisan and Míguez 2018.

[5] Pérez-Vieites, Mariño, and Míguez 2017.

[6] Pérez-Vieites and Míguez 2021.

Family of nested filters

1. **Nested particle filters (NPFs)**[4].
   - Both layers $\longrightarrow$ Sequential Monte Carlo (SMC) methods
2. **Nested hybrid filters (NHFs)**[5].
   - $\boldsymbol{\theta}$-layer $\longrightarrow$ Monte Carlo-based methods (e.g., SMC or SQMC)
   - $\boldsymbol{x}$-layer $\longrightarrow$ Gaussian techniques (e.g., EKFs or EnKFs)
3. **Nested Gaussian filters (NGFs)**[6].
   - $\boldsymbol{\theta}$-layer $\longrightarrow$ Deterministic sampling methods (e.g., UKF).
   - $\boldsymbol{x}$-layer $\longrightarrow$ Gaussian techniques (e.g., EKFs or EnKFs).

[4] Crisan and Míguez 2018.

[5] Pérez-Vieites, Mariño, and Míguez 2017.

[6] Pérez-Vieites and Míguez 2021.

Index

Adapting N_θ

Problem: Great amount of samples ($N_{\boldsymbol{\theta}} \times N_{\mathbf{x}}$) and **waste of computational effort** when they are not well chosen.

Objective: efficient allocation of computational resources

Approach: reducing automatically the number of samples, N_θ, when the performance is no longer compromised.

We studied the case for a nested Gaussian filter that implements:

- **Quadrature Kalman filter (QKF) in the θ-layer**, with $N_\theta = \alpha^{d_\theta}, \alpha > 1$.
- **Extended Kalman filters (EKFs) in the x-layer**.

The hyperparameter α will depend on t, so the number of samples is now defined as $N_{\theta,t} = \alpha_t^{d_\theta}$.

Adapting N_θ

Problem: Great amount of samples ($N_{\boldsymbol{\theta}} \times N_{\mathbf{x}}$) and **waste of computational effort** when they are not well chosen.

Objective: efficient allocation of computational resources

Approach: reducing automatically the number of samples, N_θ, when the performance is no longer compromised.

We studied the case for a nested Gaussian filter that implements:

- **Quadrature Kalman filter (QKF) in the θ-layer**, with $N_{\boldsymbol{\theta}} = \alpha^{d_{\boldsymbol{\theta}}}, \alpha > 1$.
- **Extended Kalman filters (EKFs) in the x-layer**.

The hyperparameter α will depend on t, so the number of samples is now defined as $N_{\boldsymbol{\theta},t} = \alpha_t^{d_{\boldsymbol{\theta}}}$.

Adapting $N_{\boldsymbol{\theta}}$

Problem: Great amount of samples ($N_{\boldsymbol{\theta}} \times N_{\mathbf{x}}$) and **waste of computational effort** when they are not well chosen.

Objective: efficient allocation of computational resources

Approach: reducing automatically the number of samples, N_{θ}, when the performance is no longer compromised.

We studied the case for a nested Gaussian filter that implements:

- **Quadrature Kalman filter (QKF) in the θ-layer**, with $N_{\boldsymbol{\theta}} = \alpha^{d_{\boldsymbol{\theta}}}, \alpha > 1$.
- **Extended Kalman filters (EKFs) in the x-layer**.

The hyperparameter α will depend on t, so the number of samples is now defined as $N_{\boldsymbol{\theta},t} = \alpha_t^{d_{\boldsymbol{\theta}}}$.

Adaptive reduction rule

In order to decide when to reduce $N_{\theta,t}$, we use

$$\rho_t = \frac{1}{\sum_{n=1}^{N_{\theta,t}} (\bar{s}_t^n)^2} \qquad \text{with} \qquad \bar{s}_t^n = \frac{p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)}{\sum_{n=1}^{N_{\theta,t}} p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)}$$

The statistic takes

- its **minimum value in** $\rho_t = 1$, which occurs when only one $p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)$, for $n = 1, \ldots, N_{\theta,t}$, is different from zero; and
- its **maximum value in** $\rho_t = N_{\theta,t}$, when for all $n = 1, \ldots, N_{\theta,t}$, the evaluations $p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)$ are equal.

The adaptive reduction rule:

- If $\frac{\rho_t}{N_{\theta,t}} > 1 - \epsilon$ (ρ_t is close to its maximum value),

$$N_{\theta,t+1} = (\alpha_t - 1)^{d_\theta} < N_{\theta,t}, \text{ with } N_{\theta,t+1} > N_{\min}.$$

- Otherwise, $N_{\theta,t+1} = N_{\theta,t}$.

Adaptive reduction rule

In order to decide when to reduce $N_{\boldsymbol{\theta},t}$, we use

$$\rho_t = \frac{1}{\sum_{n=1}^{N_{\boldsymbol{\theta},t}} (\bar{s}_t^n)^2} \qquad \text{with} \qquad \bar{s}_t^n = \frac{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)}{\sum_{n=1}^{N_{\boldsymbol{\theta},t}} p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)}$$

The statistic takes

- its **minimum value in** $\rho_t = 1$, which occurs when only one $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)$, for $n = 1, \ldots, N_{\boldsymbol{\theta},t}$, is different from zero; and
- its **maximum value in** $\rho_t = N_{\boldsymbol{\theta},t}$, when for all $n = 1, \ldots, N_{\boldsymbol{\theta},t}$, the evaluations $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_t^n)$ are equal.

The **adaptive reduction rule**:

- If $\frac{\rho_t}{N_{\boldsymbol{\theta},t}} > 1 - \epsilon$ (ρ_t is close to its maximum value),

$$N_{\boldsymbol{\theta},t+1} = (\alpha_t - 1)^{d_\theta} < N_{\boldsymbol{\theta},t}, \text{ with } N_{\boldsymbol{\theta},t+1} > N_{\min}.$$

- Otherwise, $N_{\boldsymbol{\theta},t+1} = N_{\boldsymbol{\theta},t}$.

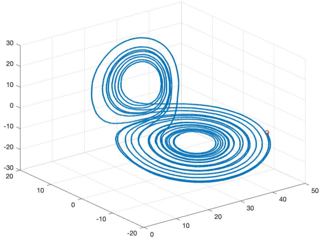
Index

The Lorenz 63 model

We consider a stochastic Lorenz 63 model, whose dynamics are described by

- the **state variables** $\boldsymbol{x}_t$ with dimension $d_x = 3$,
- the **static parameters** $\boldsymbol{\theta} = [S, R, B]^\top$ and
- the following **SDEs**

$$dx_1 = [-S(x_1 - x_2)]d\tau + \sigma dv_1,$$
$$dx_2 = [Rx_1 - x_2 - x_1 x_3]d\tau + \sigma dv_2,$$
$$dx_3 = [x_1 x_2 - Bx_3]d\tau + \sigma dv_3,$$



The Lorenz 63 model

- Applying a discretization method with step Δ, we obtain

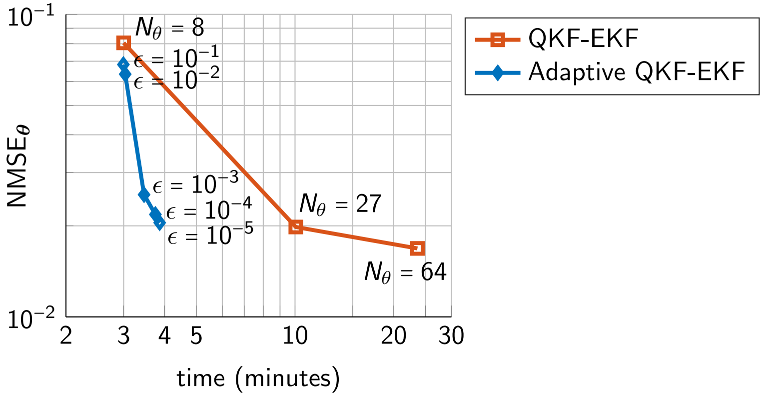
$$
\begin{aligned}
x_{1,t+1} &= x_{1,t} - \Delta S(x_{1,t} - x_{2,t}) + \sqrt{\Delta}\sigma v_{1,t},\\
x_{2,t+1} &= x_{2,t} + \Delta\left[(R - x_{3,t})x_{1,t} - x_{2,t}\right] + \sqrt{\Delta}\sigma v_{2,t},\\
x_{3,t+1} &= x_{3,t} + \Delta(x_{1,t}x_{2,t} - Bx_{3,t}) + \sqrt{\Delta}\sigma v_{3,t},
\end{aligned}
$$

- We assume linear observations of the form

$$
\boldsymbol{y}_t = k_o \begin{bmatrix} x_{1,t} \\ x_{3,t} \end{bmatrix} + \boldsymbol{r}_t,
$$

where k_o is a fixed known parameter and $\boldsymbol{r}_t \sim \mathcal{N}(\boldsymbol{r}_t|\boldsymbol{0}, \sigma_y^2 \boldsymbol{I}_2)$.

Numerical results[7]



1. QKF-EKF for different fixed $N_{\boldsymbol{\theta}} = \{8, 27, 64\}$.
2. Adaptive QKF-EKF with $N_{\boldsymbol{\theta},1} = 64$.

[7] Pérez-Vieites and Elvira 2023.

Index

Conclusions

1. The nested methodology is online and flexible. It admits different types of filtering techniques in each layer, leading to a **set of algorithms**.
2. When two learning tasks take place, it is important to have a good allocation of computational resources so that performance is improved.
3. First adaptive rule for N_{θ}, that provides an automatic reduction of the computational complexity for any given performance.

Thank you!

- Pérez-Vieites, S., & Elvira, V. (2023). *Adaptive Gaussian nested filter for parameter estimation and state tracking in dynamical systems.* In 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023).
- **Pérez-Vieites** & Míguez (2021). *Nested Gaussian filters for recursive Bayesian inference and nonlinear tracking in state space models.* Signal Processing, 189, 108295.
- **Pérez-Vieites**, Mariño & Míguez (2018). *Probabilistic scheme for joint parameter estimation and state prediction in complex dynamical systems*. Physical Review E, 98(6), 063305.
- Crisan & Míguez (2018), *Nested particle filters for online parameter estimation in discrete-time state-space Markov models.* Bernoulli, vol. 24, no. 4A, pp. 3039–3086.

sarapv.github.io

Sara Pérez Vieites