# Online Adaptation of the Number of Parameter Samples in Nested Filtering


Sara Pérez-Vieites*, Víctor Elvira†

*Department of Electrical Engineering and Automation, Aalto University, Finland.
†School of Mathematics, University of Edinburgh, United Kingdom.
**Emails**: sara.perezvieites@aalto.fi*, victor.elvira@ed.ac.uk†

Aalto-yliopisto
Aalto-universitetet
Aalto University


## Background: state-space model

We are interested in **Markov state-space dynamical systems** that can be described by the pair of equations

$$\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{v}_t, \quad (1)$$
$$\boldsymbol{y}_t = g(\boldsymbol{x}_t, \boldsymbol{\theta}) + \boldsymbol{r}_t, \quad (2)$$

where

- $\boldsymbol{x}_t \in \mathbb{R}^{d_x}$ is a **state** vector,
- $\boldsymbol{y}_t \in \mathbb{R}^{d_y}$ is the noisy **observation** vector, and
- $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$ is an **unknown parameter** vector.

We have available the pdfs $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta})$ and $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta})$.

## Background: nested filtering

**Goal**: Given prior distributions, $p(\boldsymbol{\theta}_0)$ and $p(\boldsymbol{x}_0)$, we want to compute **recursively** the joint posterior pdf of both the state and the parameters

$$p(\boldsymbol{\theta}, \boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \underbrace{p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta})}_{\boldsymbol{x}\text{-layer}} \times \underbrace{p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})}_{\boldsymbol{\theta}\text{-layer}}$$

**Approach**: We approximate the pdf of the parameters with a **sampled-based filtering method** (e.g., sequential Monte Carlo), and we run other filtering method to approximate the pdf of the state conditional on each parameter sample.

- **Initialization**: Draw $N$ particles/points $\{\boldsymbol{\theta}_0^{(i)}\} \sim p(\boldsymbol{\theta}_0)$.
- For each $t \geq 0$:
  1. Jittering step: Draw $\tilde{\boldsymbol{\theta}}_t^{(i)} \sim \kappa_N(\boldsymbol{\theta}|d\boldsymbol{\theta}_{t-1}^{(i)})$.
  2. For each $i = 1, \ldots, N$, run a **filtering method for the state**:
     a. Propagate the state: approximate
     $$\hat{p}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)})\hat{p}(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)})d\boldsymbol{x}_{t-1}. \quad (3)$$
     b. Evaluate the likelihood, $p(\boldsymbol{y}_t|\boldsymbol{x}_t, \tilde{\boldsymbol{\theta}}_t^{(i)})$, and approximate the posterior pdf,
     $$\hat{p}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \tilde{\boldsymbol{\theta}}_t^{(i)}) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t, \tilde{\boldsymbol{\theta}}_t^{(i)})\hat{p}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)}). \quad (4)$$
     c. Approximate the marginal likelihood
     $$\hat{p}(\boldsymbol{y}_t|\tilde{\boldsymbol{\theta}}_t^{(i)}, \boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t, \tilde{\boldsymbol{\theta}}_t^{(i)})\hat{p}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)})d\boldsymbol{x}_t \quad (5)$$
  3. Approximate the posterior pdf of the parameters as
  $$\hat{p}(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}) \propto \hat{p}(\boldsymbol{y}_t|\tilde{\boldsymbol{\theta}}_t^{(i)}, \boldsymbol{y}_{1:t-1})\hat{p}(\boldsymbol{\theta}|\boldsymbol{y}_{1:t-1}), \quad (6)$$
  and obtain new set of samples $\{\boldsymbol{\theta}_t^{(i)}\} \sim \hat{p}(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$.

**Family of nested methods**:

- **Nested particle filter (NPF)** [1]. Monte Carlo-based techniques.
- **Nested hybrid filter (NHF)** [2]. Monte Carlo-based in the $\boldsymbol{\theta}$-layer, and Gaussian approximations in the $\boldsymbol{x}$-layer (e.g., EKF, UKF).
- **Nested Gaussian filter (NGF)** [3]. Deterministic-sampling techniques in the $\boldsymbol{\theta}$-layer (e.g., UKF, EnKF, QKF), and other Gaussian approximations in the $\boldsymbol{x}$-layer.

## References


[1] Crisan, D., & Míguez, J. (2018). Nested particle filters for online parameter estimation in discrete-time state-space Markov models.

[2] Pérez-Vieites, S., Mariño, I. P., & Míguez, J. (2018). Probabilistic scheme for joint parameter estimation and state prediction in complex dynamical systems. Physical Review E, 98(6).

[3] Pérez-Vieites, S., & Míguez, J. (2021). Nested Gaussian filters for recursive Bayesian inference and nonlinear tracking in state space models. Signal Processing, 189, 108295.

[4] Pérez-Vieites, S., & Elvira, V. (2023, June). Adaptive Gaussian Nested Filter for Parameter Estimation and State Tracking in Dynamical Systems. In ICASSP 2023 (pp. 1-5). IEEE.


## Objective and approach

**Objective**:

- **Reduce computational complexity** of nested Gaussian filters without compromising performance.

**Approach**:

- **Reduce the number of points**, $N$, when parameters are close to convergence.
- Decision based on an **adaptive rule** that uses the **statistic** $\rho_t$.

## The statistic $\rho_t$

With $N_t$ defined as the number of samples at time $t$, the **sample quality** is assessed with:

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

- It takes its **minimum value** in $\rho_t = 1$, when only one $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)})$, is different from zero.
- It takes its **maximum value** in $\rho_t = N_t$, when for all the evaluations $p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \tilde{\boldsymbol{\theta}}_t^{(i)})$ are equal.

## Adaptive rule

We assume a NGF implementation using **cubature or quadrature rules** to approximate $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:t})$. At each time step $t$ we define

$$\{\boldsymbol{\theta}_t^{(i)}, w_t^{(i)}\}_{i=1}^{N_t} \sim \hat{p}(\boldsymbol{\theta}|\boldsymbol{y}_{1:t}), \quad \text{where} \quad N_t = \alpha_t^{d_\theta}, \ \alpha_t \in \mathbb{N}.$$

**Adaptive rule.** We set $0 < \epsilon \ll 1$ to adjust how conservative the rule is, and $1 < \alpha_{\min} < \alpha_0$, to ensure $N_t \geq 1, \forall t$.

- If $\frac{\rho_t}{N_t} < 1 - \epsilon$,
$$N_{t+1} = \alpha_{t+1}^{d_\theta} \quad \text{with} \quad \alpha_{t+1} = \max(\alpha_t - 1, \alpha_{\min}).$$
- Otherwise, $N_{t+1} = N_t$.

## Numerical Experiments

- We consider the **stochastic Lorenz 63 model** described by
$$x_{1,t} = x_{1,t-1} - \Delta S(x_{1,t-1} - x_{2,t-1}) + \sqrt{\Delta}\sigma v_{1,t},$$
$$x_{2,t} = x_{2,t-1} + \Delta[(R - x_{3,t-1})x_{1,t-1} - x_{2,t-1}] + \sqrt{\Delta}\sigma v_{2,t},$$
$$x_{3,t} = x_{3,t-1} + \Delta(x_{1,t-1}x_{2,t-1} - Bx_{3,t-1}) + \sqrt{\Delta}\sigma v_{3,t}, \quad (7)$$
where $\boldsymbol{\theta} = [S, R, B]^\top \in \mathbb{R}^3$ is the unknown static parameter vector, $\Delta$ is the integration time-step, $\{v_{i,t}\}_{i=1}^3$ are independent Wiener processes, and $\sigma > 0$.
- Comparison of NGF with a fixed $N_t = N = \alpha^{d_\theta}$, and NGF with adaptive rule, where $\alpha_0 = 4$ and $\alpha_{\min} = 2$.