Langevin-based strategies for nested particle filters

Sara Pérez-Vieites

Aalto University
Finnish Center for Artificial Intelligence (FCAI)

International Conference on Monte Carlo Methods and Applications (MCM)

31 July 2025

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

Index

Introduction

Nested particle filters

Nested filtering

Nested particle filter (NPF)

Gradient-based exploration of the parameter space

State-space model

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

$$\theta \sim p(\theta)$$
 and $\mathbf{x}_0 \sim p(\mathbf{x}_0)$ (1)

$$\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta})$$
 (2)

$$V_{\pm} \sim p(V_{\pm}|X_{\pm},\theta)$$
 (3)

State-space model

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

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

$$\theta \sim p(\theta) \text{ and } \mathbf{x}_0 \sim p(\mathbf{x}_0)$$
 (1)

$$\mathbf{x}_t \sim p(\mathbf{x}_t | \mathbf{x}_{t-1}, \boldsymbol{\theta}) \tag{2}$$

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\theta})$$
 (3)

Goal

- \longrightarrow We want to approximate the joint posterior distribution of θ and x_t , i.e., $p(x_t, \theta|y_{1:t})$.
- → For a long sequence of observations, i.e., **online**.

State-of-the-art methods

Methods for Bayesian inference of both θ and x_t :

- particle Markov chain Monte Carlo (PMCMC)¹
- sequential Monte Carlo square (SMC²)²
- nested particle filters (NPFs)³

Introduction

¹Andrieu, Doucet, and Holenstein 2010.

²Chopin, Jacob, and Papaspiliopoulos 2013.

³Crisan and Míguez 2018.

State-of-the-art methods

Methods for Bayesian inference of both θ and x_t :

- particle Markov chain Monte Carlo (PMCMC)¹
- sequential Monte Carlo square (SMC²)²
- nested particle filters (NPFs)³
 - → They can quantify the uncertainty or estimation error.
 - → They can be applied to a broad class of models.
 - They provide theoretical guarantees.

¹Andrieu, Doucet, and Holenstein 2010.

²Chopin, Jacob, and Papaspiliopoulos 2013.

³Crisan and Míguez 2018.

State-of-the-art methods

Methods for Bayesian inference of both θ and x_t :

- particle Markov chain Monte Carlo (PMCMC)¹
- sequential Monte Carlo square (SMC²)²
- nested particle filters (NPFs)³

Introduction

- → They can quantify the uncertainty or estimation error.
- → They can be applied to a broad class of models.
- They provide theoretical guarantees.
- → Both PMCMC and SMC² are batch techniques, while the NPF is a recursive method.

¹Andrieu, Doucet, and Holenstein 2010.

²Chopin, Jacob, and Papaspiliopoulos 2013.

³Crisan and Míguez 2018.

Index

Introduction

Nested particle filters

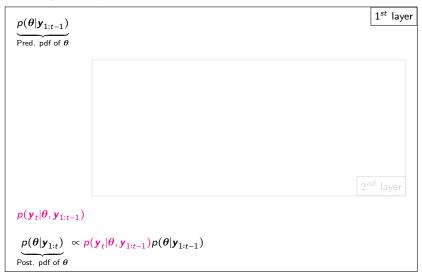
Nested filtering

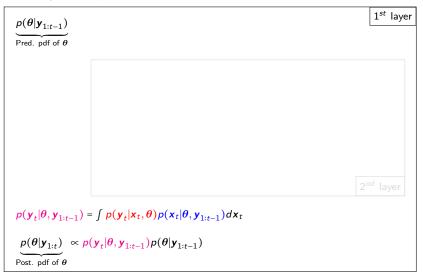
Nested particle filter (NPF)

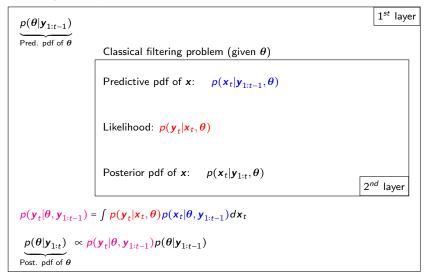
Gradient-based exploration of the parameter space

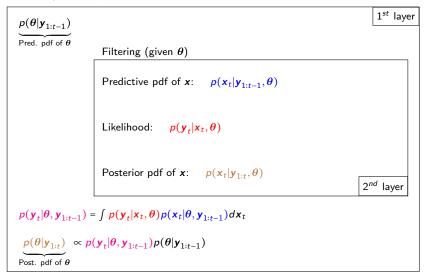
We aim at computing the joint posterior pdf $p(\theta, \mathbf{x}_t | \mathbf{y}_{1:t})$, as

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









Nested particle filter (NPF)⁴

For i = 1, ..., N:

- Jittering: Draw
$$\bar{\theta}_t^i \sim \kappa_N(d\theta|\theta_{t-1}^i)$$

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

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

Given
$$\bar{\boldsymbol{\theta}}_t^i$$
, for $j = 1, \dots, M$:

- Draw $\bar{\mathbf{x}}_{t}^{i,j} \sim p(\mathbf{x}_{t}|\mathbf{x}_{t-1}^{i,j}, \bar{\boldsymbol{\theta}}_{t}^{i})$

- Weights: $\tilde{u}_{t}^{i,j} \propto p(\mathbf{y}_{t}|\bar{\mathbf{x}}_{t}^{i,j},\bar{\boldsymbol{\theta}}_{t}^{i})$
- Resampling: for m = 1, ..., M, $\tilde{\mathbf{x}}_{t}^{i,j} = \bar{\mathbf{x}}_{t}^{i,m}$

with probability
$$u_t^{i,m} = \frac{\tilde{u}_t^{i,m}}{\sum_{j=1}^{M} \tilde{u}_t^{i,j}}$$

- Weights of $\vec{\theta}_t^i$: $\tilde{w}_t^i = \frac{1}{M} \sum_{i=1}^M \tilde{u}_t^{i,j}$
- Resampling: for l = 1, ..., N, $\{\theta_t^i, \{\mathbf{x}_t^{i,j}\}_{1 \le i \le M}\} = \{\bar{\theta}_t^l, \{\tilde{\mathbf{x}}_t^{l,j}\}_{1 \le i \le M}\}$ with prob. w_t^l , so that $p(\theta|\mathbf{y}_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\theta_i^l}(d\theta)$

Jittering

• NPF \longrightarrow jittering: $\bar{\boldsymbol{\theta}}_t^i \sim \kappa_N(d\boldsymbol{\theta}|\boldsymbol{\theta}')$, where

$$\kappa_N(d\theta|\theta') = (1 - \epsilon_N)\delta_{\theta'}(\theta) + \epsilon_N\kappa(d\theta|\theta')$$

- $0 < \epsilon_N \le \frac{1}{\sqrt{N}}$
- $\kappa(d\theta|\theta')$ is an arbitrary Markov kernel with mean θ' and finite variance, e.g., $\kappa(d\theta|\theta') = \mathcal{N}(\theta|\theta', \tilde{\sigma}^2 \mathbf{I})$, with $\tilde{\sigma}^2 < \infty$.
- Guarantees convergence to the true posterior when $N \longrightarrow \infty$

Take-aways

Advantages:

- Only framework that is *online* and *Bayesian* on heta
- Applicable to general parametric state-space models
- Asymptotic convergence guarantees

Limitations

- Covergence speed might be slow (depends on the jittering (hyper)parameters)
- This problem gets worse as the dimension of θ increases

Take-aways

Advantages:

- Only framework that is *online* and *Bayesian* on heta
- Applicable to general parametric state-space models
- Asymptotic convergence guarantees

Limitations:

- Covergence speed might be slow (depends on the jittering (hyper)parameters)
- This problem gets worse as the dimension of heta increases

Index

Introduction

Nested particle filters

Nested filtering

Nested particle filter (NPF)

Gradient-based exploration of the parameter space

Langevin nested particle filter (LNPF)

Iteratively, move the parameters towards areas of higher probability with the Unadjusted Langevin algorithm (ULA):

$$\boldsymbol{\theta}_{t,k+1}^{l} = \boldsymbol{\theta}_{t,k}^{l} + \gamma_{k} \cdot \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:t}) \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{t,k}^{l}} + \sqrt{2\gamma_{k}} \boldsymbol{v}_{k}, \tag{4}$$

where $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_{d_\theta})$, $\gamma_k > 0$ is a step size sequence.

We replace jittering by ULA

(jittering)
$$\bar{\theta}_t^i \sim \kappa_{NPF}(d\theta|\theta') = (1 - \epsilon_N)\delta_{\theta'}(\theta) + \epsilon_N \kappa_{jitter}(d\theta|\theta')$$
 (5)

(ULA)
$$\bar{\theta}_t^i \sim \kappa_{LNPF}(d\theta|\theta') = (1 - \epsilon_N)\delta_{\theta'}(\theta) + \epsilon_N \kappa_{ULA}(d\theta|\theta')$$
 (6)

*(ULA needs to adjust number of iterations and step size

Langevin nested particle filter (LNPF)

Iteratively, move the parameters towards areas of higher probability with the Unadjusted Langevin algorithm (ULA):

$$\boldsymbol{\theta}_{t,k+1}^{l} = \boldsymbol{\theta}_{t,k}^{l} + \gamma_{k} \cdot \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:t}) \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{t,k}^{l}} + \sqrt{2\gamma_{k}} \boldsymbol{v}_{k}, \tag{4}$$

where $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_{d_\theta})$, $\gamma_k > 0$ is a step size sequence.

We replace jittering by ULA:

(jittering)
$$\bar{\theta}_t^i \sim \kappa_{NPF}(d\theta|\theta') = (1 - \epsilon_N)\delta_{\theta'}(\theta) + \epsilon_N \kappa_{jitter}(d\theta|\theta')$$
 (5)

(ULA)
$$\bar{\theta}_t^i \sim \kappa_{LNPF}(d\theta|\theta') = (1 - \epsilon_N)\delta_{\theta'}(\theta) + \epsilon_N \kappa_{ULA}(d\theta|\theta')$$
 (6)

*(ULA needs to adjust number of iterations and step size)

Langevin nested particle filter (LNPF)

Iteratively, move the parameters towards areas of higher probability with the Unadjusted Langevin algorithm (ULA):

$$\boldsymbol{\theta}_{t,k+1}^{l} = \boldsymbol{\theta}_{t,k}^{l} + \gamma_{k} \cdot \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:t}) \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{t,k}^{l}} + \sqrt{2\gamma_{k}} \boldsymbol{v}_{k}, \tag{4}$$

where $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{I}_{d_\theta})$, $\gamma_k > 0$ is a step size sequence.

We replace jittering by ULA:

(jittering)
$$\bar{\theta}_t^i \sim \kappa_{NPF}(d\theta|\theta') = (1 - \epsilon_N)\delta_{\theta'}(\theta) + \epsilon_N \kappa_{jitter}(d\theta|\theta')$$
 (5)

(ULA)
$$\bar{\theta}_{t}^{i} \sim \kappa_{LNPF}(d\theta|\theta') = (1 - \epsilon_{N})\delta_{\theta'}(\theta) + \epsilon_{N} \kappa_{ULA}(d\theta|\theta')$$
 (6)

*(ULA needs to adjust number of iterations and step size)

Challenges: intractable score

The gradient of interest includes the score, that is intractable and its computation is not recursive.

$$\nabla_{\theta} \log p(\theta \mid \mathbf{y}_{1:t}) = \nabla_{\theta} \log p(\mathbf{y}_{1:t} \mid \theta) + \nabla_{\theta} \log p(\theta)$$
 (7)

→ Using Fisher's identity we have two approximations⁵

$$\mathcal{O}(N): \qquad \nabla_{\theta} \log p(\mathbf{y}_{1:t} \mid \theta) = \int \nabla_{\theta} \log p(\mathbf{y}_{1:t}, \mathbf{x}_{1:t} \mid \theta) p(\mathbf{x}_{1:t} \mid \mathbf{y}_{1:t}, \theta) d\mathbf{x}_{1:t}$$

$$\nabla_{\theta} \log p(\mathbf{y}_{1:t} \mid \theta) = \int \nabla_{\theta} \log p(\mathbf{y}_{1:t}, \mathbf{x}_{t} \mid \theta) p(\mathbf{x}_{t} \mid \mathbf{y}_{1:t}, \theta) d\mathbf{x}_{t}$$

*(This is described for a fixed θ)

Challenges: intractable score

The gradient of interest includes the score, that is intractable and its computation is not recursive.

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} \mid \mathbf{y}_{1:t}) = \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_{1:t} | \boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}) \tag{7}$$

→ Using Fisher's identity we have two approximations⁵

$$\mathcal{O}(N): \qquad \nabla_{\theta} \log p(\mathbf{y}_{1:t} \mid \theta) = \int \nabla_{\theta} \log p(\mathbf{y}_{1:t}, \mathbf{x}_{1:t} \mid \theta) p(\mathbf{x}_{1:t} \mid \mathbf{y}_{1:t}, \theta) d\mathbf{x}_{1:t}$$

$$\mathcal{O}(N^{2}): \qquad \nabla_{\theta} \log p(\mathbf{y}_{1:t} \mid \theta) = \int \nabla_{\theta} \log p(\mathbf{y}_{1:t}, \mathbf{x}_{t} \mid \theta) p(\mathbf{x}_{t} \mid \mathbf{y}_{1:t}, \theta) d\mathbf{x}_{t}$$

*(This is described for a fixed θ)

Challenges: intractable score

The gradient of interest includes the score, that is intractable and its computation is not recursive.

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} \mid \mathbf{y}_{1:t}) = \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_{1:t} | \boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}) \tag{7}$$

→ Using Fisher's identity we have two approximations⁵

$$\mathcal{O}(N): \qquad \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}) = \int \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_{1:t}, \mathbf{x}_{1:t} \mid \boldsymbol{\theta}) p(\mathbf{x}_{1:t} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}) d\mathbf{x}_{1:t}$$

$$\mathcal{O}(N^{2}): \qquad \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}) = \int \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_{1:t}, \mathbf{x}_{t} \mid \boldsymbol{\theta}) p(\mathbf{x}_{t} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}) d\mathbf{x}_{t}$$

*(This is described for a fixed θ)

Challenges: recursive approximation

 \longrightarrow Approximation used in the recursive maximum likelihood literature⁶, such that

$$\nabla_{\theta} \log p(\mathbf{y}_{1:t}|\theta)$$
 is replaced by $\nabla_{\theta} \log p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \theta)$,

and it can be computed as

$$\nabla_{\theta} \log p(\mathbf{y}_{t} \mid \mathbf{y}_{1:t-1}, \theta) \bigg|_{\theta = \theta_{t}} = \nabla_{\theta} \log p(\mathbf{y}_{1:t} | \theta) \bigg|_{\theta = \theta_{1:t}} - \nabla_{\theta} \log p(\mathbf{y}_{1:t-1} | \theta) \bigg|_{\theta = \theta_{1:t-1}}$$

*(We are assuming that process y_t is ergodic)

Challenges: recursive approximation

 \longrightarrow Approximation used in the recursive maximum likelihood literature 6 , such that

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{1:t}|\boldsymbol{\theta}) \ \text{ is replaced by } \ \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}),$$

and it can be computed as

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_t} = \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{1:t} | \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{1:t}} - \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{1:t-1} | \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{1:t-1}}$$

*(We are assuming that process y_t is ergodic)

Challenges: recursive approximation

 \longrightarrow Approximation used in the recursive maximum likelihood literature 6 , such that

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{1:t}|\boldsymbol{\theta})$$
 is replaced by $\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta})$,

and it can be computed as

$$\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}) \Bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_t} = \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{1:t} | \boldsymbol{\theta}) \Bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{1:t}} - \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{y}_{1:t-1} | \boldsymbol{\theta}) \Bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{1:t-1}}$$

*(We are assuming that process \mathbf{y}_t is ergodic)

Conclusions

- Jittering can be very inefficient, especially with larger d_{θ}
- ullet We proposed ULA updates for smarter exploration of eta space
- Challenge is approximating the score online and accurately

References

- 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.
- Chopin, N., Jacob, P. E., & Papaspiliopoulos, O. (2013). SMC²: an efficient algorithm for sequential analysis of state space models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 75(3), 397-426.
- Andrieu, Christophe, Arnaud Doucet, & Roman Holenstein (2010), Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B: Statistical Methodology 72.3: 269-342.
- Poyiadjis, G., Doucet, A., & Singh, S. S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. Biometrika, 98(1), 65-80.
- Chopin, N., & Papaspiliopoulos, O. (2020). An introduction to sequential Monte Carlo (Vol. 4). Cham, Switzerland: Springer.

Thank you!

https://sarapv.github.io/