Statistica Sinica: Supplement

Particle-based, Rapid Incremental Smoother Meets Particle Gibbs

Gabriel Cardoso, Eric Moulines, and Jimmy Olsson

Centre de Mathématiques Appliquées, Ecole polytechnique, UMR 7642, Palaiseau, France

Electrophysiology and Heart Modeling Institute (IHU-Liryc), Pessac, France

Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden

Supplementary Material

S1. Additional numerical results

S1.1 LGSSM

1 and 2 display the matrices A, B, $RR^\intercal$, and $SS^\intercal$ used for all experiments in the LGSSM model context. In 1a,2a,3a we display boxplots of bias estimates, where each estimate is obtained by averaging 10^4 independent runs of the corresponding algorithm and each box is based on 10^3 replications of this bias estimator. The `PARIS` is compared to the `PPG` for different algorithmic configurations (N, k, k_0) and for different computational budgets $C = kN$ of sizes 10^3 (1), 2.5×10^3 (2), and 5×10^3 (3). Each experiment is carried through for each of the different designs $k_0 = \lfloor 2^{-1} k \rfloor$, $k_0 = \lfloor (3/4) C/N \rfloor$,

and $k_0 = k - 1$ of the burn-in.

/	1	2	3	4	5
1	-0.4193	0.00182	0.00183	0.00184	0.00185
2	0.2145	0.63952	0.63953	0.63954	0.63955
3	0.3449	0.60202	0.60203	0.60204	0.60205
4	0.2572	-0.26932	-0.26933	-0.26934	-0.26935
5	0.7505	-0.36332	-0.36333	-0.36334	-0.36335

/	1	2	3	4	5
1	-0.2078	0.27752	0.27753	0.27754	0.27755
2	0.0984	0.45172	0.45173	0.45174	0.45175
3	0.7050	-0.04502	-0.04503	-0.04504	-0.04505
4	0.1684	-0.15152	-0.15153	-0.15154	-0.15155
5	-0.0320	0.50612	0.50613	0.50614	0.50615

Table 1: The A (left) and B (right) matrices in the LGSSM.

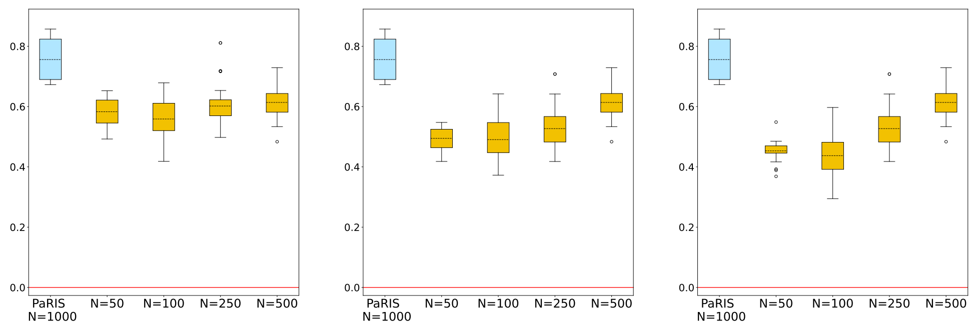
/	1	2	3	4	5
1	0.0026	-0.00062	-0.00063	-0.00064	-0.00065
2	-0.0004	0.00122	0.00123	0.00124	0.00125
3	-0.0001	-0.00062	-0.00063	-0.00064	-0.00065
4	0.0007	0.00012	0.00013	0.00014	0.00015
5	-0.0006	0.00282	0.00283	0.00284	0.00285

/	1	2	3	4	5
1	0.0157	-0.00072	-0.00073	-0.00074	-0.00075
2	0.0014	0.00072	0.00073	0.00074	0.00075
3	-0.0027	0.00592	0.00593	0.00594	0.00595
4	0.0064	-0.01052	-0.01053	-0.01054	-0.01055
5	-0.0007	0.02072	0.02073	0.02074	0.02075

Table 2: The covariance matrices $RR^\intercal$ (left) and $SS^\intercal$ (right) for the state and measurement noises, respectively, in the LGSSM.

S1.2 Stochastic volatility

In this section we repeat the same experiments in S1.1 in the context of the StoVol model described in 5. 4–6 display boxplots of bias estimates for the `PARIS` and the `PPG` for different algorithmic configurations (N, k, k_0) and different computational budgets $C = kN$ of sizes 10^2 (4), 5×10^2 (5), and 10^3 (6). The bias of each algorithm is estimated by averaging 10^3 independent runs of the same, and each box is based on 10^3 independent replications of this bias estimator. Again, in each plot, the `PARIS` and `PPG` share the same computational budget (regardless configuration of the `PPG`).

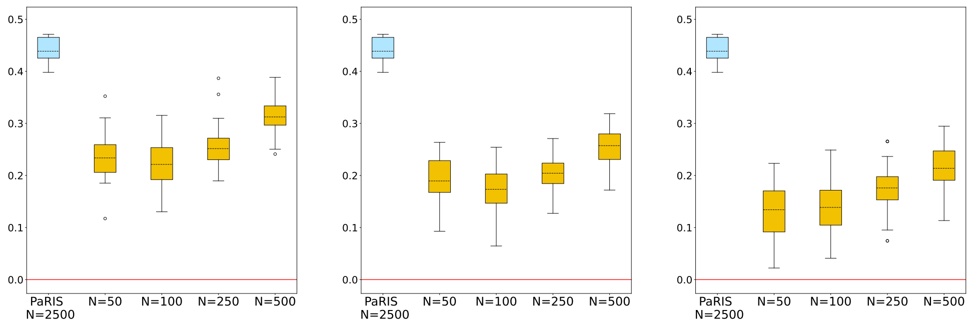


(a) $k_0 = \lfloor 2^{-1}C/N \rfloor$ (b) $k_0 = \lfloor (3/4)C/N \rfloor$ (c) $k_0 = k - 1$

Figure 1: PARIS and PPG outputs for the LGSSM with $C = 10^3$ and different designs of the burn-in k_0.

Choice of (N, k, k_0)**.** Designing the configuration (N, k, k_0) is challenging, since the upper bound $\kappa_{N,n}$ on the mixing rate is known to be conservative. As clear from 4–6, the best configuration also depends on C; indeed, we see that for a smaller budget it is better to let the particle sample size N be large. Nevertheless, for more generous budgets it seems to be better to use a large number k of iterations at the expense of N.

Concerning the burn-in parameter k_0, the choice depends mainly on the bias–variance trade-off. In applications where minimising the bias is important one would choose $k_0 = k - 1$, which gives the smallest possible bias. Otherwise, a trade-off that provides an improvement in bias at the cost of an increase in MSE over the PARIS by only a factor of 2 is to choose $k_0 = \lfloor k/2 \rfloor$; recall the discussion in 4.2.

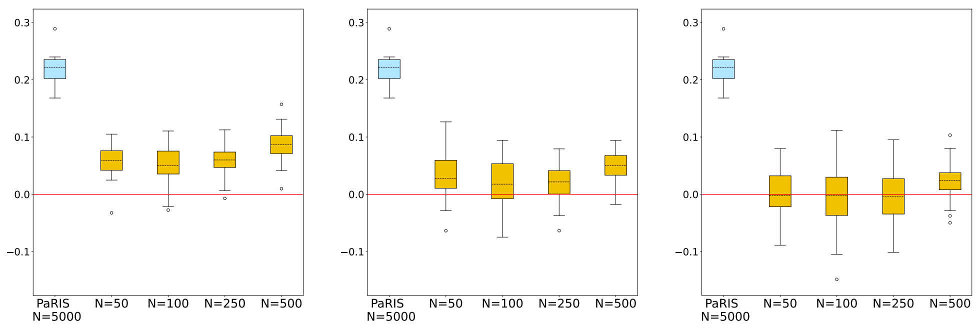


(a) $k_0 = \lfloor 2^{-1} C/N \rfloor$ (b) $k_0 = \lfloor (3/4) C/N \rfloor$ (c) $k_0 = k - 1$

Figure 2: `PARIS` and `PPG` outputs for the LGSSM with $C = 2.5 \times 10^3$ and different designs of the burn-in k_0.

Comparison with the Rhee–Glynn-type estimator of [3]

We now compare the proposed `PPG` estimator with the unbiased Rhee–Glynn-type smoothing estimator $\mathrm{H}_{k_0:k,N}$ defined in [3, Eq. 2], where the parameter k_0 is the burn-in phase length, k the minimum number of Gibbs iterations, and N the number of particles used in the coupled conditional particle filter. This estimator is based on the *coupled conditional particle filter* with ancestor sampling proposed in [3]; see 4 for details. Since the number of particles used in the algorithm is itself a random variable, we first perform 3×10^3 independent runs of the same and report the average *meeting time* (*i.e.*, number of iterations of 4 until the conditional paths $\zeta_{0:n}$ and $\zeta'_{0:n}$ become identical) for three different choices of the hyperparameters in 3. We deduce from 3 that the average total number of particles generated



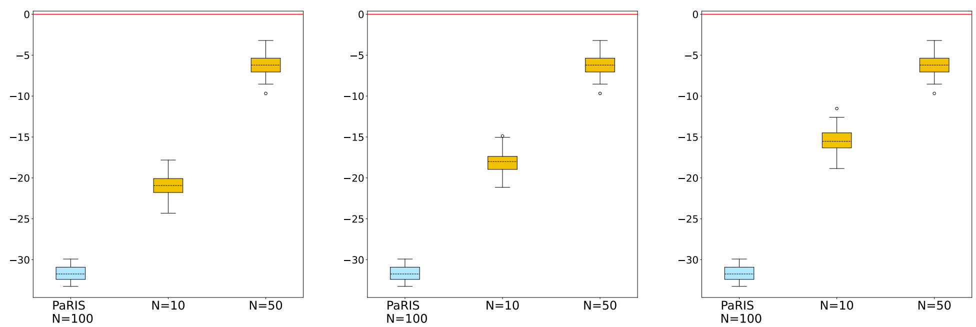
(a) $k_0 = \lfloor 2^{-1} C/N \rfloor$ (b) $k_0 = \lfloor (3/4) C/N \rfloor$ (c) $k_0 = k - 1$

Figure 3: PARIS and PPG outputs for the LGSSM with $C = 5 \times 10^3$ and different designs of the burn-in k_0.

N	k_0	k	Meeting time
100	5	10	30.4
250	2	4	12.6
500	1	2	7.1

Table 3: Coupled conditional particle filter meeting times for three different configurations with $Nk = 10^3$.

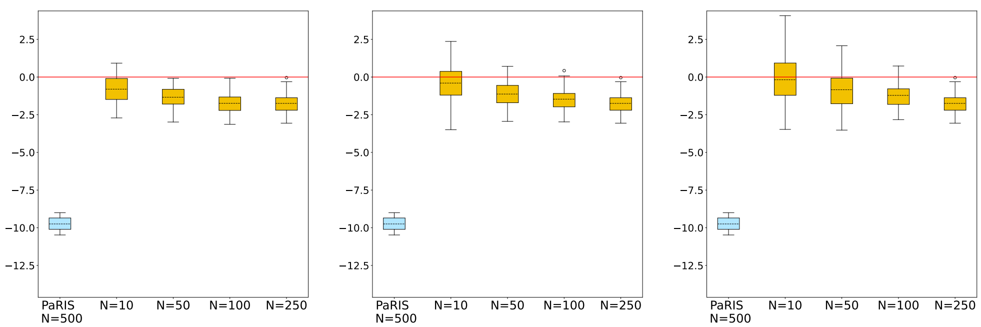
is about 3×10^3. Therefore, we compare the Rhee–Glynn estimator induced by the coupled conditional particle filter with the PPG estimator with $(N, k_0, k) = (10, 150, 300)$. 7 shows histograms of estimates produced using the Rhee–Glynn-type procedure, for the three different configurations, along with histograms of the estimates produced by the PPG. Each histogram is based on 3×10^3 independent replications. We find that the variance and empirical bias of the Rhee–Glynn-type estimator is about 10 and 20 times larger, respectively, than for the PPG for the same computational effort.



(a) $k_0 = \lfloor 2^{-1} C/N \rfloor$ (b) $k_0 = \lfloor (3/4) C/N \rfloor$ (c) $k_0 = k - 1$

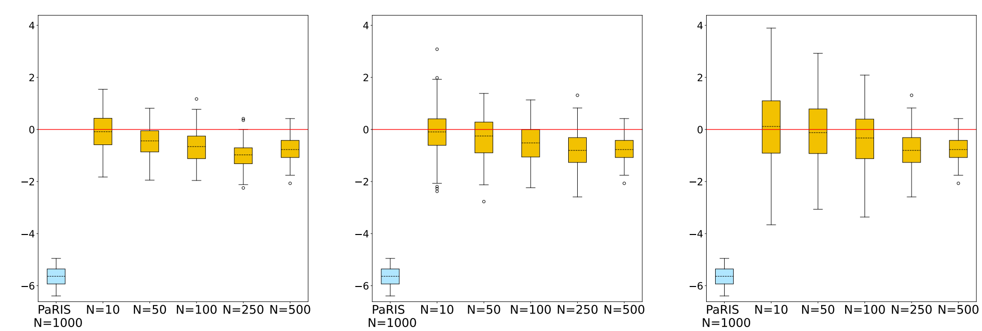
Figure 4: `PARIS` and `PPG` outputs for the stovol model with $C = 10^2$ and different designs on the burn-in k_0.

Another way of obtaining Rhee–Glynn-type smoothing estimator would be to consider the coupling of the conditional backward sampling particle filter, as proposed in [5]. In the case of the bootstrap particle filter, the conditional particle filter with backward sampling is probabilistically equivalent to the conditional particle filter with ancestor sampling. Furthermore, [5, Section 7] also show that for $n = 10^3$, both the conditional particle filter with backward sampling and the conditional particle filter with ancestor sampling have similar performance. Thus, we expect the results in this section to translate to the estimators proposed in [5].



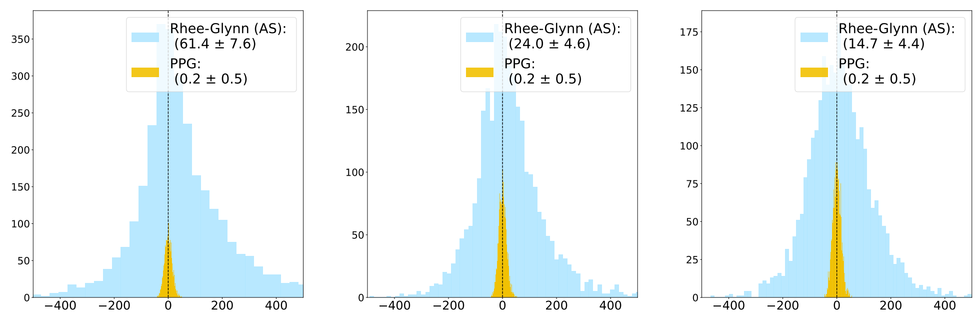
(a) $k_0 = \lfloor 2^{-1}C/N \rfloor$ (b) $k_0 = \lfloor (3/4)C/N \rfloor$ (c) $k_0 = k - 1$

Figure 5: `PARIS` and `PPG` outputs for the stovol model with $C = 5 \times 10^2$ and different designs of the burn-in k_0.



(a) $k_0 = \lfloor 2^{-1}C/N \rfloor$ (b) $k_0 = \lfloor (3/4)C/N \rfloor$ (c) $k_0 = k - 1$

Figure 6: `PARIS` and `PPG` outputs for the stovol model with $C = 10^3$ and different designs of the burn-in k_0.



(a) $(N, k) = (100, 10)$ (b) $(N, k) = (250, 4)$ (c) $(N, k) = (500, 2)$

Figure 7: Histograms of estimates produced using the Rhee–Glynn-type smoothing estimator of [3] for three different configurations and the `PPG` estimator with $(N, k_0, k) = (10, 150, 300)$. Each box is based on 3000 independent replications. The plot also provides the corresponding 95% coverage asymptotic confidence intervals.

S2. Algorithms

The following section provides pseudocode for the algorithms discussed in 3, namely: the original PARIS algorithm (1) proposed in [7], the conditional PARIS update (2), and the PPG (3). In addition, we provide a pseudocode for the coupled conditional conditional particle filter with *ancestor sampling* (4), being the key ingredient of the unbiased Rhee–Glynn-type estimator proposed in [3] against which the PPG is benchmarked in S1.2. Note that the conditional PARIS update described in 2 differs somewhat from that described in 3 in the way the underlying conditional dual process $\{\boldsymbol{\xi}_m\}_{m \in \mathbb{N}}$ is propagated; more precisely, in 2, each conditional dual process update $\boldsymbol{\xi}_{m+1} \sim \boldsymbol{M}_m \langle \zeta_{m+1} \rangle (\boldsymbol{\xi}_m, \cdot)$, where the value of ζ_{m+1} is inserted into a randomly chosen position in $\boldsymbol{\xi}_{m+1}$ (whereas the remaining elements of $\boldsymbol{\xi}_{m+1}$ are sampled independently from $\Phi_m(\mu(\boldsymbol{\xi}_m))$) is replaced by deterministic assignment of ζ_{m+1} to ξ_{m+1}^N. Of course, this change has no impact as long as we are interested in integrating functions that are permutation invariant with respect to the produced many-body systems, which is the case throughout our work. Still, as this derandomization technique simplifies somewhat the implementation of the PPG, we have chosen to include it in our pseudocode.

Data: $\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$

Result: $\{(\xi_{n+1}^i, \beta_{n+1}^i)\}_{i=1}^N$

for $i \leftarrow 1$ **to** N **do**

draw $I_{n+1}^i \sim \mathsf{cat}(\{g_n(\xi_n^\ell)\}_{\ell=1}^N)$;

draw $\xi_{n+1}^i \sim M_n(\xi_n^{I_{n+1}^i}, \cdot)$;

for $j \leftarrow 1$ **to** M **do**

draw $J_{n+1}^{(i,j)} \sim \mathsf{cat}(\{q_n(\xi_n^\ell, \xi_{n+1}^i)\}_{\ell=1}^N)$

end

set $\beta_{n+1}^i \leftarrow \frac{1}{M} \sum_{j=1}^M \left(\beta_n^{J_{n+1}^{(i,j)}} + \tilde{h}_n(\xi_n^{J_{n+1}^{(i,j)}}, \xi_{n+1}^i) \right)$;

end

Algorithm 1: One update of the `PARIS`.

Coupling algorithms. 4 provides a more detailed description of (the predictive variant of) the coupled conditional particle filter proposed in [3, Algorithm 1], and we focus here on the version of this algorithm where the iteratively produced particle paths underlying the resulting estimator are generated by means of ancestor sampling [6]. If $\{\omega_\ell\}_{\ell=1}^N$ and $\{\omega'_\ell\}_{\ell=1}^N$ are possibly unnormalized event probabilities, we denote by $\mathsf{M}(\{\omega_\ell\}_{\ell=1}^N, \{\omega'_\ell\}_{\ell=1}^N)$ the *maximal coupling* between the distributions $\mathsf{cat}(\{\omega_\ell\}_{\ell=1}^N)$ and $\mathsf{cat}(\{\omega'_\ell\}_{\ell=1}^N)$. In our implementations, we used the maximum coupling given in [4, Algorithm 2]. In order to couple two conditional particle filters, we assume, following [3, Algorithm 1], that for every $m \in \mathbb{N}$ we are able to simulate a random variable ε_m, defined on some measurable space $(\mathsf{S}_m, \mathcal{S}_m)$

Data: $\boldsymbol{\upsilon}_n$, ζ_{n+1}

Result: $\boldsymbol{\upsilon}_{n+1}$

for $i \leftarrow 1$ **to** $N-1$ **do**

  draw $I^i_{m+1} \sim \mathsf{cat}(\{g_m(\xi^\ell_{m|m})\}^N_{\ell=1})$;

  draw $\xi^i_{m+1|m+1} \sim M_m(\xi^{I^i_{m+1}}_{m|m}, \cdot)$;

end

set $\xi^N_{m+1|m+1} \leftarrow \zeta_{m+1}$;

for $i \leftarrow 1$ **to** N **do**

  for $j \leftarrow 1$ **to** M **do**

    draw $J^{(i,j)}_{m+1} \sim \mathsf{cat}(\{q_m(\xi^\ell_{m|m}, \xi^i_{m+1|m+1})\}^N_{\ell=1})$

  end

  set $\beta^i_{m+1} \leftarrow \frac{1}{M}\sum_{j=1}^M \left(\beta_m^{J^{(i,j)}_{m+1}} + \tilde{h}_m(\xi^{J^{(i,j)}_{m+1}}_{m|m}, \xi^i_{m+1|m+1})\right)$;

  set $\xi^i_{0:m+1|m+1} \leftarrow (\xi^{J^{(i,1)}_{m+1}}_{0:m|m}, \xi^i_{m+1|m+1})$;

end

set $\boldsymbol{\upsilon}_{n+1} \leftarrow ((\xi^1_{0:n+1|n+1}, \beta^1_{n+1}), \dots, (\xi^N_{0:n+1|n+1}, \beta^N_{n+1}))$;

Algorithm 2: One conditional `PARIS` update, expressed in a short form as "$\boldsymbol{\upsilon}_{n+1} \leftarrow \mathsf{CondPaRIS}(\boldsymbol{\upsilon}_n, \zeta_{n+1})$".

and distributed according $\mu_m \in \mathsf{M}_1(\mathcal{S}_m)$, such that there exists some measurable function ϕ on $(\mathsf{X}_m \times \mathsf{S}_m, \mathcal{X}_m \otimes \mathcal{S}_m)$ such that for every $x_m \in \mathsf{X}_m$, $\mu_m \circ \phi_m^{-1}(x_m, \cdot)$ (the pushforward of μ_m through $\phi_m(x_m, \cdot)$) equals $M_m(x_m, \cdot)$.

Data: $\zeta_{0:n}$

Result: $\boldsymbol{\upsilon}_n$, $\zeta'_{0:n}$

draw $(\xi^1_{0|0}, \ldots, \xi^{N-1}_{0|0}) \sim \eta_0^{\otimes(N-1)}$;

set $\xi^N_{0|0} \leftarrow \zeta_0$;

set $\boldsymbol{\beta}_0 \leftarrow (0, \ldots, 0)$;

for $m \leftarrow 0$ **to** $n-1$ **do**

  run $((\xi^1_{m+1|m+1}, \beta^1_{m+1}), \ldots, (\xi^N_{m+1|m+1}, \beta^N_{m+1})) \leftarrow$

  $\mathsf{CondPaRIS}((\xi^1_{m|m}, \beta^1_m), \ldots, (\xi^N_{m|m}, \beta^N_m), \zeta_{m+1})$;

end

set $\boldsymbol{\upsilon}_n \leftarrow ((\xi^1_{n|n}, \beta^1_n), \ldots, (\xi^N_{n|n}, \beta^N_n))$;

draw $J \sim \mathsf{cat}(\{1\}_{\ell=1}^N)$;

set $\zeta'_{0:n} \leftarrow \xi^J_{0:n|n}$;

Algorithm 3: One iteration of the Parisian particle Gibbs (`PPG`)

Data: $\zeta_{0:n}$, $\tilde{\zeta}_{0:n}$

Result: $\zeta'_{0:n}$, $\tilde{\zeta}'_{0:n}$

set $(\xi_0^1, \ldots, \xi_0^{N-1}) \sim \eta_0^{\otimes(N-1)}$;

set $(\tilde{\xi}_0^1, \ldots, \tilde{\xi}_0^{N-1}) \leftarrow (\xi_0^1, \ldots, \xi_0^{N-1})$;

set $(\xi_0^N, \tilde{\xi}_0^N) \leftarrow (\zeta_0, \tilde{\zeta}_0)$;

for $m \leftarrow 0$ **to** $n-1$ **do**

  for $i \leftarrow 1$ **to** $N-1$ **do**

    draw $(I_{m+1}^i, \tilde{I}_{m+1}^i) \sim \mathsf{M}(\{g_m(\xi_m^\ell)\}_{\ell=1}^N, \{g_m(\tilde{\xi}_m^\ell)\}_{\ell=1}^N)$;

  end

  draw $(I_{m+1}^N, \tilde{I}_{m+1}^N) \sim \mathsf{M}(\{q_m(\xi_m^\ell, \zeta_{m+1})\}_{\ell=1}^N, \{q_m(\tilde{\xi}_m^\ell, \tilde{\zeta}_{m+1})\}_{\ell=1}^N)$;

  for $i \leftarrow 1$ **to** N **do**

    draw $\varepsilon_m \sim \mu_m$;

    set $(\xi_{m+1}^i, \tilde{\xi}_{m+1}^i) \leftarrow (\phi_m(\xi_m^{I_{m+1}^i}, \varepsilon_m), \phi_m(\tilde{\xi}_m^{\tilde{I}_{m+1}^i}, \varepsilon_m))$;

  end

end

draw $J_n \sim \mathsf{cat}(\{1\}_{\ell=1}^N)$;

set $\tilde{J}_n \leftarrow J_n$;

set $(\zeta_n, \tilde{\zeta}_n) \leftarrow (\xi_n^{J_n}, \tilde{\xi}_n^{\tilde{J}_n})$;

for $m \leftarrow n-1$ **to** 0 **do**

  set $(J_m, \tilde{J}_m) \leftarrow (I_{m+1}^{J_{m+1}}, \tilde{I}_{m+1}^{\tilde{J}_{m+1}})$;

  set $(\zeta_m, \tilde{\zeta}_m) \leftarrow (\xi_m^{J_m}, \tilde{\xi}_m^{\tilde{J}_m})$;

end

Algorithm 4: Coupled conditional particle filters [3].

S3. Additional proofs

S3.1 Proof of 2

First, note that, by definitions (3.1) and (3.2),

$$
\begin{aligned}
H_n(\boldsymbol{x}_{0:n}) &\coloneqq \int \mathbb{S}_n(\boldsymbol{x}_{0:n}, \mathrm{d}\boldsymbol{y}_n)\, \mu(\boldsymbol{x}_{0:n|n}) h \\
&= \int \cdots \int \left(\frac{1}{N} \sum_{j_n=1}^{N} h(x_{0:n-1|n}^{j_n}, x_n^{j_n}) \right) \\
&\quad \times \prod_{m=0}^{n-1} \prod_{i_{m+1}=1}^{N} \int \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}),
\end{aligned}
$$

where $x_{0:-1|0}^{i} = \emptyset$ for all $i \in [\![1, N]\!]$ by convention. We will show that for every $k \in [\![0, n]\!]$, $H_{k,n} \equiv H_n$, where

$$
H_{k,n}(\boldsymbol{x}_{0:n}) \coloneqq \frac{1}{N} \sum_{j_n=1}^{N} \cdots \sum_{j_k=1}^{N} \prod_{\ell=k}^{n-1} \frac{q_\ell(x_\ell^{j_\ell}, x_{\ell+1}^{j_{\ell+1}})}{\sum_{j'_\ell=1}^{N} q_\ell(x_\ell^{j'_\ell}, x_{\ell+1}^{j_{\ell+1}})} a_{k,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \ldots, x_n^{j_n})
$$

with

$$
\begin{aligned}
&a_{k,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \ldots, x_n^{j_n}) \\
&= \int \prod_{m=0}^{k-1} \prod_{i_{m+1}=1}^{N} \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}) h(x_{0:k-1|k}^{j_k}, x_k^{j_k}, \ldots, x_n^{j_n}).
\end{aligned}
$$

Since, by convention, $\prod_{\ell=n}^{n-1} \ldots = 1$, $H_{n,n}(\boldsymbol{x}_{0:n}) = N^{-1} \sum_{j_n=1}^{N} a_{n,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{n-1}, x_n^{j_n})$, and we note that $H_n \equiv H_{n,n}$. We now show that $H_{k,n} \equiv H_{k-1,n}$ for every

$k \in [\![1, n]\!]$; for this purpose, note that

$$
\begin{aligned}
&a_{k,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \ldots, x_n^{j_n}) \\
&\qquad = \int \prod_{m=0}^{k-2} \prod_{i_{m+1}=1}^{N} \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}) \\
&\times \int \prod_{i_k=1}^{N} \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{i_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{i_k})} \delta_{x_{0:k-1|k-1}^{j_{k-1}}}(\mathrm{d}x_{0:k-1|k}^{i_k})\, h(x_{0:k-1|k}^{j_k}, x_k^{j_k}, \ldots, x_n^{j_n}),
\end{aligned}
$$

and since $x_{0:k-1|k-1}^{j_{k-1}} = (x_{0:k-2|k-1}^{j_{k-1}}, x_{k-1}^{j_{k-1}})$, it holds that

$$
\begin{aligned}
&\int \prod_{i_k=1}^{N} \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{i_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{i_k})} \delta_{x_{0:k-1|k-1}^{j_{k-1}}}(\mathrm{d}x_{0:k-1|k}^{i_k})\, h(x_{0:k-1|k}^{j_k}, x_k^{j_k}, \ldots, x_n^{j_n}) \\
&\qquad = \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} h(x_{0:k-2|k-1}^{j_{k-1}}, x_{k-1}^{j_{k-1}}, x_k^{j_k}, \ldots, x_n^{j_n}).
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
&a_{k,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \ldots, x_n^{j_n}) \\
&\qquad = \int \prod_{m=0}^{k-2} \prod_{i_{m+1}=1}^{N} \sum_{j_m=1}^{N} \frac{q_m(x_m^{j_m}, x_{m+1}^{i_{m+1}})}{\sum_{j'_m=1}^{N} q_m(x_m^{j'_m}, x_{m+1}^{i_{m+1}})} \delta_{x_{0:m|m}^{j_m}}(\mathrm{d}x_{0:m|m+1}^{i_{m+1}}) \\
&\qquad\quad \times \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} h(x_{0:k-2|k-1}^{j_{k-1}}, x_{k-1}^{j_{k-1}}, x_k^{j_k}, \ldots, x_n^{j_n}).
\end{aligned}
$$

Now, changing the order of summation with respect to j_{k-1} and integration on the right hand side of the previous display yields

$$
\begin{aligned}
&a_{k,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-1}, x_k^{j_k}, \ldots, x_n^{j_n}) \\
&\qquad = \sum_{j_{k-1}=1}^{N} \frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} a_{k-1,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-2}, x_{k-1}^{j_{k-1}}, \ldots, x_n^{j_n}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&H_{k,n}(\boldsymbol{x}_{0:n}) \\
&= \frac{1}{N}\sum_{j_n=1}^{N}\cdots\sum_{j_k=1}^{N}\prod_{\ell=k}^{n-1}\frac{q_\ell(x_\ell^{j_\ell}, x_{\ell+1}^{j_{\ell+1}})}{\sum_{j'_\ell=1}^{N} q_\ell(x_\ell^{j'_\ell}, x_{\ell+1}^{j_{\ell+1}})} \\
&\quad \times \sum_{j_{k-1}=1}^{N}\frac{q_{k-1}(x_{k-1}^{j_{k-1}}, x_k^{j_k})}{\sum_{j'_{k-1}=1}^{N} q_{k-1}(x_{k-1}^{j'_{k-1}}, x_k^{j_k})} a_{k-1,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-2}, x_{k-1}^{j_{k-1}}, \ldots, x_n^{j_n}) \\
&= \frac{1}{N}\sum_{j_n=1}^{N}\cdots\sum_{j_{k-1}=1}^{N}\prod_{\ell=k-1}^{n-1}\frac{q_\ell(x_\ell^{j_\ell}, x_{\ell+1}^{j_{\ell+1}})}{\sum_{j'_\ell=1}^{N} q_\ell(x_\ell^{j'_\ell}, x_{\ell+1}^{j_{\ell+1}})} a_{k-1,n}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-2}, x_{k-1}^{j_{k-1}}, \ldots, x_n^{j_n}) \\
&= H_{k-1,n}(\boldsymbol{x}_{0:n}),
\end{aligned}
$$

which establishes the recursion. Therefore, $H_n \equiv H_{0,n}$ and we may now conclude the proof by noting that $\mathbb{B}_n h \equiv H_{0,n}$.

S3.2 Proof of 5

In order to establish 5 we will prove the following more general result, of which 5 is a direct consequence.

Proposition 1. *For every $n \in \mathbb{N}$ and $M \in \mathbb{N}^*$ there exist $\mathsf{c}_n > 0$ and $\mathsf{d}_n > 0$ such that for every $N \in \mathbb{N}^*$, $z_{0:n} \in \mathsf{X}_{0:n}$, $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$, and*

$\varepsilon > 0$,

$$\int \mathbb{C}_n \mathbb{S}_n(z_{0:n}, \mathrm{d}\boldsymbol{b}_n) \\ \times \mathbb{1}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\{b_n^i f_n(x_{n|n}^i) + \tilde{f}_n(x_{n|n}^i)\} - \eta_n\langle z_{0:n}\rangle(f_n B_n\langle z_{0:n-1}\rangle h_n + \tilde{f}_n)\right| \geq \varepsilon\right\} \\ \leq \mathsf{c}_n \exp\left(-\frac{\mathsf{d}_n N\varepsilon^2}{2\mathsf{\kappa}_n^2}\right),$$

where

$$\mathsf{\kappa}_n := \|f_n\|_\infty \sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty + \|\tilde{f}_n\|_\infty. \tag{S3.1}$$

To prove 1 we need the following technical lemma.

Lemma 1. *For every $n \in \mathbb{N}$, $(f_{n+1}, \tilde{f}_{n+1}) \in \mathsf{F}(\mathcal{X}_{n+1})^2$, $z_{0:n+1} \in \mathsf{X}_{0:n+1}$, and $N \in \mathbb{N}^*$,*

$$\gamma_{n+1}\langle z_{0:n+1}\rangle(f_{n+1}B_{n+1}\langle z_{0:n}\rangle h_{n+1} + \tilde{f}_{n+1}) \\ = \left(1 - \frac{1}{N}\right)\gamma_n\langle z_{0:n}\rangle\{Q_n f_{n+1} B_n\langle z_{0:n-1}\rangle h_n + Q_n(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\} \\ + \frac{1}{N}\gamma_n\langle z_{0:n}\rangle g_n\left(f_{n+1}(z_{n+1})B_{n+1}\langle z_{0:n}\rangle h_{n+1}(z_{n+1}) + \tilde{f}_{n+1}(z_{n+1})\right).$$

Proof. Since 2 holds also for the Feynman–Kac model with a frozen path, we obtain

$$\gamma_{n+1}\langle z_{0:n+1}\rangle(f_{n+1}B_{n+1}\langle z_{0:n}\rangle h_{n+1} + \tilde{f}_{n+1}) \\ = \gamma_n\langle z_{0:n}\rangle\{Q_n\langle z_{n+1}\rangle f_{n+1}B_n\langle z_{0:n}\rangle h_n + Q_n\langle z_{n+1}\rangle(\tilde{h}_n f_{n+1} + \tilde{f}_{n+1})\}.$$

Thus, the proof is concluded by noting that for every $x_n \in \mathsf{X}_n$ and $h \in \mathsf{F}(\mathcal{X}_{n:n+1})$,

$$Q_n\langle z_{n+1}\rangle h(x_n) = \left(1 - \frac{1}{N}\right) Q_n h(x_n) + \frac{1}{N} g(x_n) h(x_n, z_{n+1}).$$

□

Finally, before proceeding to the proof of 1, we introduce the law of the PARIS evolving conditionally on a frozen path $z = \{z_m\}_{m\in\mathbb{N}}$. Define, for $m \in \mathbb{N}$ and $z_{m+1} \in \mathsf{X}_{m+1}$,

$$\boldsymbol{P}_m\langle z_{m+1}\rangle : \mathbf{Y}_m \times \boldsymbol{\mathcal{Y}}_{m+1} \ni (\boldsymbol{y}_m, A) \mapsto \int \boldsymbol{M}_m\langle z_{m+1}\rangle(\boldsymbol{x}_{m|m}, \mathrm{d}\boldsymbol{x}_{m+1})\, \boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, A).$$

For any given initial distribution $\boldsymbol{\psi}_0 \in \mathsf{M}_1(\boldsymbol{\mathcal{Y}}_0)$, let $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\psi}_0}$ be the distribution of the canonical Markov chain induced by the Markov kernels $\{\boldsymbol{P}_m\langle z_{m+1}\rangle\}_{m\in\mathbb{N}}$ and the initial distribution $\boldsymbol{\psi}_0$. By abuse of notation we write $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}$ instead of $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\psi}_0[\boldsymbol{\eta}_0\langle z_0\rangle]}$, where the extension $\boldsymbol{\psi}_0[\boldsymbol{\eta}_0]$ is defined in 6.3.

Proof of 1. We proceed by forward induction over n. Let the σ-fields $\tilde{\mathcal{F}}_n$ and $\mathcal{F}_n$ be defined as in the proof of 3, but for the conditional PARIS dual process. Then, under the law $\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}$, reusing (6.11),

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left[\, \beta^1_n f_n(\xi^1_n) + \tilde{f}_n(\xi^1_n) \mid \tilde{\mathcal{F}}_{n-1}\right] \\
&= \mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left[\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left[\, \beta^1_n \mid \mathcal{F}_n\right]\ f_n(\xi^1_n) + \tilde{f}_n(\xi^1_n) \mid \tilde{\mathcal{F}}_{n-1}\right] \\
&= \mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left[\ f_n(\xi^1_n) \sum_{\ell=1}^{N} \frac{q_{n-1}(\xi^\ell_{n-1}, \xi^1_n)}{\sum_{\ell'=1}^N q_{n-1}(\xi^{\ell'}_{n-1}, \xi^1_n)} \left(\beta^\ell_{n-1} + \tilde{h}_{n-1}(\xi^\ell_{n-1}, \xi^1_n)\right) + \tilde{f}_n(\xi^1_n) \mid \tilde{\mathcal{F}}_{n-1}\right].
\end{aligned}$$

Using (2.6), we get

$$\mathbb{E}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left[\beta^1_n f_n(\xi^1_n) + \tilde{f}_n(\xi^1_n) \mid \tilde{\mathcal{F}}_{n-1}\right]$$
$$= \left(1 - \frac{1}{N}\right) \frac{\sum_{\ell=1}^N \{\beta^\ell_{n-1} Q_{n-1} f_n(\xi^\ell_{n-1}) + Q_{n-1}(\tilde{h}_{n-1} f_n + \tilde{f}_n)(\xi^\ell_{n-1})\}}{\sum_{\ell'=1}^N g_{n-1}(\xi^{\ell'}_{n-1})}$$
$$+ \frac{1}{N}\left(f_n(z_n) \sum_{\ell=1}^N \frac{q_{n-1}(\xi^\ell_{n-1}, z_n)}{\sum_{\ell'=1}^N q_{n-1}(\xi^{\ell'}_{n-1}, z_n)} \left(\beta^\ell_{n-1} + \tilde{h}_n(\xi^\ell_{n-1}, z_n)\right) + \tilde{f}_n(z_n)\right). \tag{S3.2}$$

In order to apply the induction hypothesis to each term on the right-hand side of the previous identity, note that

$$B_n\langle z_{0:n-1}\rangle h_n(z_n) = \frac{\eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot, z_n)\{B_{n-1}\langle z_{0:n-2}\rangle h_{n-1}(\cdot) + \tilde{h}_{n-1}(\cdot, z_n)\}]}{\eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot, z_n)]}.$$

Therefore, using 1 and noting that $\gamma_n\langle z_{0:n}\rangle \mathbb{1}_{\mathsf{X}_n} / \gamma_{n-1}\langle z_{0:n}\rangle \mathbb{1}_{\mathsf{X}_{n-1}} = \eta_{n-1}\langle z_{0:n-1}\rangle g_{n-1}$ yields

$$\eta_n\langle z_{0:n}\rangle(f_n B_n\langle z_{0:n-1}\rangle h_n + \tilde{f}_n) = \frac{1}{N}\left(f_n(z_n) B_n\langle z_{0:n-1}\rangle h_n(z_n) + \tilde{f}_n(z_n)\right)$$
$$+ \left(1 - \frac{1}{N}\right) \frac{\eta_{n-1}\langle z_{0:n-1}\rangle\{Q_{n-1} f_n B_{n-1}\langle z_{0:n-2}\rangle h_n + Q_{n-1}(\tilde{h}_{n-1} f_n + \tilde{f}_n)\}}{\eta_{n-1}\langle z_{0:n-1}\rangle g_{n-1}}. \tag{S3.3}$$

By combining (S3.2) with (S3.3), we decompose the error according to

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\{\beta_n^i f_n(\xi_{n|n}^i)+\tilde{f}_n(\xi_{n|n}^i)\}-\eta_n\langle z_{0:n}\rangle(f_n B_n\langle z_{0:n-1}\rangle h_n+\tilde{f}_n)\\
&=\frac{1}{N}\sum_{i=1}^{N}\{\beta_n^i f_n(\xi_{n|n}^i)+\tilde{f}_n(\xi_{n|n}^i)\}-\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left[\,\beta_n^1 f_n(\xi_n^1)+\tilde{f}_n(\xi_n^1)\mid\tilde{\mathcal{F}}_{n-1}\right]\\
&\quad+\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left[\beta_n^1 f_n(\xi_n^1)+\tilde{f}_n(\xi_n^1)\mid\tilde{\mathcal{F}}_{n-1}\right]-\eta_n\langle z_{0:n}\rangle(f_n B_n\langle z_{0:n-1}\rangle h_n+\tilde{f}_n)\\
&=\mathrm{I}_N^{(1)}+\left(1-\frac{1}{N}\right)\mathrm{I}_N^{(2)}+\frac{1}{N}\,\mathrm{I}_N^{(3)},
\end{aligned}
\tag{S3.4}
$$

where

$$
\begin{aligned}
\mathrm{I}_N^{(1)}&\coloneqq\frac{1}{N}\sum_{i=1}^{N}\{\beta_n^i f_n(\xi_n^i)+\tilde{f}_n(\xi_n^i)\}-\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left[\beta_n^1 f_n(\xi_n^1)+\tilde{f}_n(\xi_n^1)\mid\tilde{\mathcal{F}}_{n-1}\right],\\
\mathrm{I}_N^{(2)}&\coloneqq\frac{\sum_{\ell=1}^{N}\{\beta_{n-1}^\ell Q_{n-1}f_n(\xi_{n-1}^\ell)+Q_{n-1}(\tilde{h}_{n-1}f_n+\tilde{f}_n)(\xi_{n-1}^\ell)\}}{\sum_{\ell'=1}^{N}g_{n-1}(\xi_{n-1}^{\ell'})}\\
&\qquad-\frac{\eta_{n-1}\langle z_{0:n-1}\rangle\{Q_{n-1}f_n B_n\langle z_{0:n-1}\rangle h_n+Q_{n-1}(\tilde{h}_{n-1}f_n+\tilde{f}_n)\}}{\eta_{n-1}\langle z_{0:n-1}\rangle g_{n-1}},
\end{aligned}
\tag{S3.5}
$$

and

$$
\begin{aligned}
\mathrm{I}_N^{(3)}&\coloneqq f_n(z_n)\sum_{\ell=1}^{N}\frac{q_{n-1}(\xi_{n-1}^\ell,z_n)}{\sum_{\ell'=1}^{N}q_{n-1}(\xi_{n-1}^{\ell'},z_n)}\left(\beta_{n-1}^\ell+\tilde{h}_{n-1}(\xi_{n-1}^\ell,z_n)\right)\\
&\quad-f_n(z_n)\frac{\eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot,z_n)\{B_{n-1}\langle z_{0:n-2}\rangle h_{n-1}(\cdot)+\tilde{h}_{n-1}(\cdot,z_n)\}]}{\eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot,z_n)]}.
\end{aligned}
\tag{S3.6}
$$

The proof is now completed by treating the terms $\mathrm{I}_N^{(1)}$, $\mathrm{I}_N^{(2)}$, and $\mathrm{I}_N^{(3)}$ separately, using Hoeffding's inequality and its generalisation in [2, Lemma 4].

Choose $\varepsilon > 0$; then, by Hoeffding's inequality,

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\,\mathrm{I}_N^{(1)}\,\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{1}{2}\frac{\varepsilon^2}{\mathsf{\kappa}_n^2}N\right). \tag{S3.7}$$

To treat $\mathrm{I}_N^{(2)}$, we apply the induction hypothesis to the numerator and denominator, each normalized by $1/N$, yielding, since $\|Q_{n-1}h\|_\infty \leq \bar{\tau}_{n-1}\|h\|_\infty$ for all $h \in \mathsf{F}(\mathcal{X}_{n-1} \otimes \mathcal{X}_n)$,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\Bigg(\Bigg|\frac{1}{N}\sum_{\ell=1}^{N}\{\beta_{n-1}^{\ell}Q_{n-1}f_n(\xi_{n-1}^{\ell}) + Q_{n-1}(\tilde{h}_{n-1}f_n + \tilde{f}_n)(\xi_{n-1}^{\ell})\}\\ -\eta_{n-1}\langle z_{0:n-1}\rangle\{Q_{n-1}f_nB_n\langle z_{0:n-1}\rangle h_n + Q_{n-1}(\tilde{h}_{n-1}f_n + \tilde{f}_n)\}\Bigg| \geq \varepsilon\Bigg)\\ \leq \mathsf{c}_{n-1}\exp\left(-\mathsf{d}_{n-1}\frac{\varepsilon^2}{\bar{\tau}_{n-1}^2\mathsf{\kappa}_n^2}N\right)\end{aligned}$$

and

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\frac{1}{N}\sum_{\ell=1}^{N}g_{n-1}(\xi_{n-1}^{\ell}) - \eta_{n-1}\langle z_{0:n-1}\rangle g_{n-1}\right| \geq \varepsilon\right) \leq \mathsf{c}_{n-1}\exp\left(-\mathsf{d}_{n-1}\frac{\varepsilon^2}{\bar{\tau}_{n-1}^2}N\right).$$

Combining the previous two bounds with the generalised Hoeffding inequality in [2, Lemma 4] yields, using also the bounds

$$\frac{\sum_{\ell=1}^{N}\{\beta_{n-1}^{\ell}Q_{n-1}f_n(\xi_{n-1}^{\ell}) + Q_{n-1}(\tilde{h}_{n-1}f_n + \tilde{f}_n)(\xi_{n-1}^{\ell})\}}{\sum_{\ell'=1}^{N}g_{n-1}(\xi_{n-1}^{\ell'})} \leq \mathsf{\kappa}_n$$

and $\eta_{n-1}\langle z_{0:n-1}\rangle g_{n-1} \geq \underline{\tau}_{n-1}$, the inequality

$$\mathbb{P}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}\left(\left|\,\mathrm{I}_N^{(2)}\,\right| \geq \varepsilon\right) \leq \mathsf{c}_{n-1}\exp\left(-\mathsf{d}_{n-1}\frac{\underline{\tau}_{n-1}^2\varepsilon^2}{\bar{\tau}_{n-1}^2\mathsf{\kappa}_n^2}N\right). \tag{S3.8}$$

The last term $\mathrm{I}_N^{(3)}$ is treated along similar lines; indeed, by the induction hypothesis, since $\|q_{n-1}\|_\infty \leq \bar{\tau}_{n-1}\bar{\sigma}_{n-1}$,

$$\begin{aligned}\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\Bigg(\Bigg|\frac{1}{N}\sum_{\ell=1}^N q_{n-1}(\xi^\ell_{n-1}, z_n)\left(\beta^\ell_{n-1} + \tilde{h}_{n-1}(\xi^\ell_{n-1}, z_n)\right) \\ - \eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot, z_n)\{B_{n-1}\langle z_{0:n-1}\rangle h_{n-1}(\cdot) + \tilde{h}_{n-1}(\cdot, z_n)\}]\Bigg| \geq \varepsilon\Bigg) \\ \leq \mathsf{c}_{n-1}\exp\left(-\mathsf{d}_{n-1}\left(\frac{\varepsilon}{\bar{\tau}_{n-1}\bar{\sigma}_{n-1}\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty}\right)^2 N\right)\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left(\left|\frac{1}{N}\sum_{\ell=1}^N q_{n-1}(\xi^\ell_{n-1}, z_n) - \eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot, z_n)]\right| \geq \varepsilon\right) \\ \leq \mathsf{c}_{n-1}\exp\left(-\mathsf{d}_{n-1}\left(\frac{\varepsilon}{\bar{\tau}_{n-1}\bar{\sigma}_{n-1}}\right)^2 N\right).\end{aligned}$$

Thus, since

$$\sum_{\ell=1}^N \frac{q_{n-1}(\xi^\ell_{n-1}, z_n)}{\sum_{\ell'=1}^N q_{n-1}(\xi^{\ell'}_{n-1}, z_n)}\left(\beta^\ell_{n-1} + \tilde{h}_{n-1}(\xi^\ell_{n-1}, z_n)\right) \leq \sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty$$

and $\eta_{n-1}\langle z_{0:n-1}\rangle[q_{n-1}(\cdot, z_n)] \geq \underline{\tau}_{n-1}$, the generalised Hoeffding inequality provides

$$\mathbb{P}^{\boldsymbol{P},z}_{\boldsymbol{\eta}_0}\left(|\,\mathrm{I}^{(3)}_N\,| \geq \varepsilon\right) \leq \mathsf{c}_{n-1}\exp\left(-\mathsf{d}_{n-1}\left(\frac{\underline{\tau}_{n-1}\varepsilon}{2\bar{\tau}_{n-1}\bar{\sigma}_{n-1}\|f_n\|_\infty\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty}\right)^2 N\right). \tag{S3.9}$$

Finally, combining the bounds (S3.7–S3.9) completes the proof. □

S3.3 Proof of 3

The statement of 3 is implied by the following more general result, which we will prove below.

Proposition 2. *For every $n \in \mathbb{N}$, $M \in \mathbb{N}^*$, $N \in \mathbb{N}^*$, $z_{0:n} \in \mathsf{X}_{0:n}$, $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$, and $p \geq 2$, it holds that*

$$\int \mathbb{C}_n \mathbb{S}_n(z_{0:n}, \mathrm{d}\boldsymbol{b}_n) \left| \frac{1}{N} \sum_{i=1}^{N} \{b_n^i f_n(x_{n|n}^i) + \tilde{f}_n(x_{n|n}^i)\} - \eta_n \langle z_{0:n} \rangle (f_n B_n \langle z_{0:n-1} \rangle h_n + \tilde{f}_n) \right|^p \leq \mathsf{c}_n (p/\mathsf{d}_n)^{p/2} N^{-p/2} \mathsf{\kappa}_n^p,$$

where $\mathsf{c}_n > 0$, $\mathsf{d}_n > 0$ and $\mathsf{\kappa}_n$ are defined in 1 and (S3.1), respectively.

Before proving 2, we establish the following result.

Lemma 2. *Let X be an $\mathbb{R}^d$-valued random variable, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying $\mathbb{P}(|X| \geq t) \leq c \exp(-t^2/(2\sigma^2))$ for every $t \geq 0$ and some $c > 0$ and $\sigma > 0$. Then for every $p \geq 2$ it holds that $\mathbb{E}[|X|^p] \leq c p^{p/2} \sigma^p$.*

Proof. Using Fubini's theorem and the change of variable formula,

$$\mathbb{E}\left[|X|^p\right] = \int_0^\infty p t^{p-1} \mathbb{P}(|X| \geq t) \, \mathrm{d}t = c p 2^{p/2-1} \sigma^p \Gamma(p/2),$$

where Γ is the Gamma function. It remains to apply the bound $\Gamma(p/2) \leq (p/2)^{p/2-1}$ (see [1]), which holds for $p \geq 2$ by [2, Theorem 1.5]. □

Proof of 2. By combining 1 and 2 we obtain

$$N \int \mathbb{C}_n \mathbb{S}_n(z_{0:n}, \mathrm{d}\boldsymbol{b}_n) \left| \frac{1}{N} \sum\nolimits_{i=1}^{N} \{ b_n^i f_n(x_{n|n}^i) + \tilde{f}_n(x_{n|n}^i) \} - \eta_n \langle z_{0:n} \rangle (f_n B_n \langle z_{0:n-1} \rangle h_n + \tilde{f}_n) \right|^2$$
$$\leq \mathsf{c}_n (p/\mathsf{d}_n)^{p/2} N^{-p/2} \left(\|f_n\|_\infty \sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty + \|\tilde{f}_n\|_\infty \right)^p,$$

which was to be established. □

S3.4 Proof of 4

Like previously, we establish 4 via a more general result, namely the following.

Proposition 3. *For every $n \in \mathbb{N}$, the exists $\bar{\mathsf{c}}_n^{bias} < \infty$ such that for every $M \in \mathbb{N}^*$, $N \in \mathbb{N}^*$, $z_{0:n} \in \mathsf{X}_{0:n}$, and $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$,*

$$\left| \int \mathbb{C}_n \mathbb{S}_n(z_{0:n}, \mathrm{d}\boldsymbol{b}_n) \frac{1}{N} \sum_{i=1}^{N} \{ b_n^i f_n(x_{n|n}^i) + \tilde{f}_n(x_{n|n}^i) \} - \eta_n \langle z_{0:n} \rangle (f_n B_n \langle z_{0:n-1} \rangle h_n + \tilde{f}_n) \right|$$
$$\leq \bar{\mathsf{c}}_n^{bias} \kappa_n N^{-1},$$

where κ_n is defined in (S3.1).

We preface the proof of 3 by a technical lemma providing a bound on the bias of ratios of random variables.

Lemma 3. *Let α and β be (possibly dependent) random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and such that $\mathbb{E}[\alpha^2] < \infty$ and $\mathbb{E}[\beta^2] < \infty$.*

Moreover, assume that there exist $c > 0$ *and* $d > 0$ *such that* $|\alpha/\beta| \leq c$, $\mathbb{P}$*-a.s.,* $|a/b| \leq c$, $\mathbb{E}[(\alpha - a)^2] \leq c^2 d^2$, *and* $\mathbb{E}[(\beta - b)^2] \leq d^2$. *Then*

$$|\mathbb{E}[\alpha/\beta] - a/b| \leq 2c(d/b)^2 + c|\mathbb{E}[\beta - b]|/|b| + |\mathbb{E}[\alpha - a]|/|b|. \tag{S3.10}$$

Proof. Using the identity

$$\mathbb{E}[\alpha/\beta] - a/b = \mathbb{E}[(\alpha/\beta)(b-\beta)^2]/b^2 + \mathbb{E}[(\alpha - a)(b - \beta)]/b^2 + a\mathbb{E}[b - \beta]/b^2 + \mathbb{E}[\alpha - a]/b,$$

the claim is established by applying the Cauchy–Schwarz inequality and the assumptions of the lemma according to

$$\begin{aligned} &|\mathbb{E}[\alpha/\beta] - a/b| \\ &\quad \leq c\mathbb{E}[(\beta - b)^2]/b^2 + \{\mathbb{E}[(\alpha - a)^2]\mathbb{E}[(\beta - b)^2]\}^{1/2}/b^2 + |a||\mathbb{E}[\beta - b]|/b^2 + |\mathbb{E}[\alpha - a]|/b^2 \\ &\quad \leq 2c(d/b)^2 + c|\mathbb{E}[\beta - b]|/|b| + |\mathbb{E}[\alpha - a]|/|b|. \end{aligned}$$

□

Proof of 4. We proceed by induction and assume that the claim holds true for $n-1$. Reusing the error decomposition (S3.4), it is enough to bound the expectations of the terms $\mathrm{I}_N^{(2)}$ and $\mathrm{I}_N^{(3)}$ given in (S3.5) and (S3.6), respectively (since $\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[\mathrm{I}_N^{(1)}] = 0$). This will be done using the induction hypothesis, 3, and 2. More precisely, to bound the expectation of $\mathrm{I}_N^{(2)}$, we use 3 with

$\alpha \leftarrow \alpha_n$, $\beta \leftarrow \beta_n$, $a \leftarrow a_n$, and $b \leftarrow b_n$, where

$$\alpha_n := \frac{1}{N}\sum_{\ell=1}^{N}\{\beta_{n-1}^{\ell} Q_{n-1} f_n(\xi_{n-1}^{\ell}) + Q_{n-1}(\tilde{h}_{n-1} f_n + \tilde{f}_n)(\xi_{n-1}^{\ell})\}, \quad \beta_n := \frac{1}{N}\sum_{\ell=1}^{N} g_{n-1}(\xi_{n-1}^{\ell}),$$

$$a_n := \eta_{n-1}\langle z_{0:n-1}\rangle\{Q_{n-1} f_n B_n\langle z_{0:n-1}\rangle h_n + Q_{n-1}(\tilde{h}_{n-1} f_n + \tilde{f}_n)\}, \quad b_n := \eta_{n-1}\langle z_{0:n-1}\rangle g_{n-1}.$$

For this purpose, note that $|\alpha_n/\beta_n| \leq \kappa_n$ and $|a_n/b_n| \leq \kappa_n$, where κ_n is defined in (S3.1). On the other hand, using 2 (applied with $p = 2$), we obtain

$$\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[(\alpha_n - a_n)^2] \leq d_n^2 \kappa_n^2 \quad \text{and} \quad \mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[(\beta_n - b_n)^2] \leq d_n^2,$$

where $d_n^2 := \mathsf{c}_n \bar{\tau}_{n-1}^2/(\mathsf{d}_n N)$. Using the induction assumption, we get

$$|\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[\alpha_n] - a_n| \leq \bar{\mathsf{c}}_{n-1}^{bias} N^{-1} \bar{\tau}_{n-1} \kappa_n \quad \text{and} \quad |\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[\beta_n] - b_n| \leq \bar{\mathsf{c}}_{n-1}^{bias} N^{-1} \bar{\tau}_{n-1}.$$

Hence, the conditions of 3 are satisfied and we deduce that

$$|\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[\mathrm{I}_N^{(2)}]| = |\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[\alpha_n/\beta_n] - a_n/b_n| \leq 2\kappa_n \frac{\mathsf{c}_n}{\mathsf{d}_n N}\frac{\bar{\tau}_{n-1}^2}{\underline{\tau}_{n-1}^2} + 2\bar{\mathsf{c}}_{n-1}^{bias}\kappa_n \frac{\bar{\tau}_{n-1}}{\underline{\tau}_{n-1} N}.$$

The bound on $|\mathbb{E}_{\boldsymbol{\eta}_0}^{\boldsymbol{P},z}[\mathrm{I}_N^{(2)}]|$ is obtained along the same lines. □

S3.5 Proof of 6

We first consider the bias, which can be bounded according to

$$\begin{aligned}\left|\mathbb{E}_{\xi}[\Pi_{(k_0,k),N}(f)] - \eta_{0:n} h_n\right| &\leq (k-k_0)^{-1}\sum_{\ell=k_0+1}^{k}|\mathbb{E}_{\xi}\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n} h_n| \\ &\leq (k-k_0)^{-1} N^{-1}\mathsf{c}_n^{bias}\left(\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty\right)\sum_{\ell=k_0+1}^{k}\kappa_{N,n}^{\ell},\end{aligned}$$

from which the bound (4.7) follows immediately.

We turn to the MSE. Using the decomposition

$$\mathbb{E}_\xi[(\Pi_{(k_0,k),N}(f) - \eta_{0:n}h_n)^2] \leq (k-k_0)^{-2} \left\{ \sum_{\ell=k_0+1}^{k} \mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n)^2] \right.$$
$$\left. + 2 \sum_{\ell=k_0+1}^{k} \sum_{j=\ell+1}^{k} \mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n)(\mu(\boldsymbol{\beta}_n[j])(\mathrm{id}) - \eta_{0:n}h_n)] \right\},$$

the MSE bound in 2 implies that

$$\sum_{\ell=k_0+1}^{k} \mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n)^2] \leq \mathsf{c}_n^{mse} \left(\sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty \right)^2 N^{-1}(k-k_0).$$

Moreover, using the covariance bound in 2, we deduce that

$$\sum_{\ell=k_0+1}^{k} \sum_{j=\ell+1}^{k} \mathbb{E}_\xi[(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n)(\mu(\boldsymbol{\beta}_n[j])(\mathrm{id}) - \eta_{0:n}h_n)]$$
$$\leq \mathsf{c}_n^{cov} \left(\sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty \right)^2 N^{-3/2} \left(\sum_{\ell=k_0+1}^{k} \sum_{j=\ell+1}^{k} \kappa_{N,n}^{(j-\ell)} \right).$$

Thus, the proof is concluded by noting that $\sum_{\ell=k_0+1}^{k} \sum_{j=\ell+1}^{k} \kappa_{N,n}^{(j-\ell)} \leq (k - k_0)/(1-\kappa_{N,n})$.

Bibliography

[1] G. D. Anderson and S.-L. Qiu. A monotonicity property of the gamma function. *Proc. Amer. Math. Soc.*, 125(11):3355–3362, 1997.

[2] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte

Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.*, 21(6):1201–2145, 2011.

[3] P. E. Jacob, F. Lindsten, and T. B. Schön. Smoothing with couplings of conditional particle filters. *J. Am. Statist. Assoc.*, 115(530):721–729, 2020.

[4] P. E. Jacob, J. O'Leary, and Y. F. Atchadé. Unbiased Markov chain Monte Carlo methods with couplings. *J. Roy. Statist. Soc. B*, 82(3):543–600, 2020.

[5] A. Lee, S. S. Singh, and M. Vihola. Coupled conditional backward sampling particle filter. 48(5):3066–3089, 2020.

[6] F. Lindsten, M. I. Jordan, and T. B. Schön. Particle Gibbs with ancestor sampling. *J. Mach. Learn. Res.*, 15(1):2145–2184, 2014.

[7] J. Olsson and J. Westerborn. Efficient particle-based online smoothing in general hidden Markov models: The PaRIS algorithm. *Bernoulli*, 23(3):1951–1996, 2017.