Smoothing with Couplings of Conditional Particle Filters: supplementary materials

Before proving Theorem 3.1 of the main document, we introduce an intermediate result on the probability of the chains meeting at the next step, irrespective of their current states. For convenience, we first recall the three assumptions of the main document.

Assumption 1 *The measurement density of the model is bounded from above: there exists* $\bar{g} < \infty$ *such that, for all* $y \in \mathbb{Y}$ *and* $x \in \mathbb{X}$, $g(y|x) \leq \bar{g}$.

Assumption 2 *The resampling probability matrix* P*, with rows summing to* $w^{1:N}$ *and columns summing to* $\tilde{w}^{1:N}$*, is such that, for all* $i \in \{1, \dots, N\}$, $P^{ii} \geq w^i \tilde{w}^i$*. Furthermore, if* $w^{1:N} = \tilde{w}^{1:N}$*, then* P *is a diagonal matrix with entries given by* $w^{1:N}$.

Assumption 3 *Let* $(X^{(n)})_{n \geq 0}$ *be a Markov chain generated by the conditional particle filter and started from* π_0*, and* h *a test function of interest. Then* $\mathbb{E}\left[h(X^{(n)})\right] \xrightarrow[n\to\infty]{} \pi(h)$*. Furthermore, there exists* $\delta > 0$, $n_0 < \infty$ *and* $C < \infty$ *such that, for all* $n \geq n_0$, $\mathbb{E}\left[h(X^{(n)})^{2+\delta}\right] \leq C$.

1 Intermediate result on the meeting probability

The result provides a lower-bound on the probability of meeting in one step, for coupled chains generated by the coupled conditional particle filter (CCPF) kernel.

Lemma 1.1 *Let* $N \geq 2$ *and* $T \geq 1$ *be fixed. Under Assumptions 1 and 2, there exists* $\varepsilon > 0$*, depending on* N *and* T*, such that*

$$\forall X \in \mathbb{X}^{T+1}, \quad \forall \tilde{X} \in \mathbb{X}^{T+1}, \quad \mathbb{P}(X' = \tilde{X}'|X, \tilde{X}) \geq \varepsilon,$$

where $(X', \tilde{X}') \sim$ *CCPF*$((X, \tilde{X}), \cdot)$. *Furthermore, if* $X = \tilde{X}$, *then* $X' = \tilde{X}'$ *almost surely.*

The constant ε depends on N and T, and on the coupled resampling scheme being used. Lemma 1.1 can be used, together with the coupling inequality (Lindvall, 2002), to prove the ergodicity of the conditional particle filter kernel, which is akin to the approach of Chopin and Singh (2015). The coupling inequality states that the total variation distance between $X^{(n)}$ and $\tilde{X}^{(n-1)}$ is less than $2\mathbb{P}(\tau > n)$, where τ is the meeting time. By assuming $\tilde{X}^{(0)} \sim \pi$, $\tilde{X}^{(n)}$ follows π at each step n, and we obtain a bound for the total variation distance between $X^{(n)}$ and π. Using Lemma 1.1, we can bound the probability $\mathbb{P}(\tau > n)$ from above by $(1-\varepsilon)^n$, as in the proof of Theorem 3.1 below. This implies that the computational cost of the proposed estimator has a finite expectation for all $N \geq 2$ and $T \geq 1$.

Proof of Lemma 1.1. We write $\mathbb{P}_{x_{0:t}, \tilde{x}_{0:t}}$ and $\mathbb{E}_{x_{0:t}, \tilde{x}_{0:t}}$ for the conditional probability and expectation, respectively, with respect to the law of the particles generated by the CCPF procedure conditionally on the reference trajectories up to time t, $(x_{0:t}, \tilde{x}_{0:t})$. Furthermore, let $\mathcal{F}_t$ denote the filtrations generated by the CCPF at time t. We denote by $x^k_{0:t}$, for $k \in 1:N$, the surviving trajectories at time t. Let $I_t \subseteq 1:N-1$ be the set of common particles at time t defined by $I_t = \{j \in 1:N-1 : x^j_{0:t} = \tilde{x}^j_{0:t}\}$. The meeting probability can then be bounded by:

$$\begin{aligned}
\mathbb{P}_{x_{0:T}, \tilde{x}_{0:T}}(x'_{0:T} = \tilde{x}'_{0:T}) &= \mathbb{E}_{x_{0:T}, \tilde{x}_{0:T}}\left[\mathbb{1}\left(x^{b_T}_{0:T} = \tilde{x}^{\tilde{b}_T}_{0:T}\right)\right] \geq \sum_{k=1}^{N-1} \mathbb{E}_{x_{0:T}, \tilde{x}_{0:T}}[\mathbb{1}(k \in I_T)\, P^{kk}_T] \\
&= (N-1)\mathbb{E}_{x_{0:T}, \tilde{x}_{0:T}}[\mathbb{1}(1 \in I_T)\, P^{11}_T] \geq \frac{N-1}{(N\bar{g})^2}\mathbb{E}_{x_{0:T}, \tilde{x}_{0:T}}[\mathbb{1}(1 \in I_T)\, g_T(x^1_T) g_T(\tilde{x}^1_T)], \quad (1)
\end{aligned}$$

where we have used Assumptions 1 and 2.

Now, let $\psi_t : \mathbb{X}^t \mapsto \mathbb{R}_+$ and consider

$$\mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}(1 \in I_t)\, \psi_t(x^1_{0:t})\psi_t(\tilde{x}^1_{0:t})] = \mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}(1 \in I_t)\, \psi_t(x^1_{0:t})^2], \tag{2}$$

since the two trajectories agree on $\{1 \in I_t\}$. We have

$$\mathbb{1}(1 \in I_t) \geq \sum_{k=1}^{N-1} \mathbb{1}(k \in I_{t-1})\, \mathbb{1}\left(a^1_{t-1} = \tilde{a}^1_{t-1} = k\right), \tag{3}$$

and thus

$$\begin{aligned}
&\mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}(1 \in I_t)\, \psi_t(x^1_{0:t})^2] \\
&\quad \geq \mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\sum_{k=1}^{N-1} \mathbb{1}(k \in I_{t-1})\, \mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}\left(a^1_{t-1} = \tilde{a}^1_{t-1} = k\right) \psi_t(x^1_{0:t})^2 \mid \mathcal{F}_{t-1}]] \\
&\quad = (N-1)\mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}(1 \in I_{t-1})\, \mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}\left(a^1_{t-1} = \tilde{a}^1_{t-1} = 1\right) \psi_t(x^1_{0:t})^2 \mid \mathcal{F}_{t-1}]].
\end{aligned} \tag{4}$$

The inner conditional expectation can be computed as

$$\begin{aligned}
&\mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}\left(a^1_{t-1} = \tilde{a}^1_{t-1} = 1\right) \psi_t(x^1_{0:t})^2 \mid \mathcal{F}_{t-1}] \\
&\qquad = \sum_{k,\ell=1}^{N} P^{k\ell}_{t-1}\mathbb{1}(k = \ell = 1) \int \psi_t((x^k_{0:t-1}, x_t))^2 f(dx_t | x^k_{t-1}) \\
&\qquad = P^{11}_{t-1} \int \psi_t((x^1_{0:t-1}, x_t))^2 f(dx_t | x^1_{t-1}) \\
&\qquad \geq \frac{g_{t-1}(x^1_{t-1})g_{t-1}(\tilde{x}^1_{t-1})}{(N\bar{g})^2} \left(\int \psi_t((x^1_{0:t-1}, x_t)) f(dx_t | x^1_{t-1}) \right)^2,
\end{aligned} \tag{5}$$

where we have again used Assumptions 1 and 2. Note that this expression is independent of the final states of the reference trajectories, $(x_t, \tilde{x}_t)$, which can thus be dropped from the conditioning. Furthermore, on $\{1 \in I_{t-1}\}$ it holds that $x^1_{0:t-1} = \tilde{x}^1_{0:t-1}$ and therefore, combining Eqs. (2)–(5) we get

$$
\begin{aligned}
&\mathbb{E}_{x_{0:t},\tilde{x}_{0:t}}[\mathbb{1}(1 \in I_t)\, \psi_t(x^1_{0:t})\psi_t(\tilde{x}^1_{0:t})] \\
&\qquad \geq \frac{(N-1)}{(N\bar{g})^2}\mathbb{E}_{x_{0:t-1},\tilde{x}_{0:t-1}}\Big[\mathbb{1}(1 \in I_{t-1})\, g_{t-1}(x^1_{t-1}) \int \psi_t((x^1_{0:t-1}, x_t)) f(dx_t|x^1_{t-1}) \\
&\qquad\qquad \times g_{t-1}(\tilde{x}^1_{t-1}) \int \psi_t((\tilde{x}^1_{0:t-1}, x_t)) f(dx_t|\tilde{x}^1_{t-1})\Big]. \qquad (6)
\end{aligned}
$$

Thus, if we define for $t = 1, \ldots, T-1$, $\psi_t(x_{0:t}) = g_t(x_t) \int \psi_{t+1}(x_{0:t+1}) f(dx_{t+1}|x_t)$, and $\psi_T(x_{0:T}) = g_T(x_T)$, it follows that

$$
\begin{aligned}
\mathbb{P}_{x_{0:T},\tilde{x}_{0:T}}(x'_{0:T} = \tilde{x}'_{0:T}) &\geq \frac{(N-1)^{\mathsf{T}}}{(N\bar{g})^{2T}}\mathbb{E}_{x_0,\tilde{x}_0}[\mathbb{1}(1 \in I_1)\, \psi_1(x^1_1)\psi_1(\tilde{x}^1_1)] \\
&= \frac{(N-1)^{\mathsf{T}}}{(N\bar{g})^{2T}}\mathbb{E}_{x_0,\tilde{x}_0}[\psi_1(x^1_1)^2] \geq \frac{(N-1)^{\mathsf{T}}}{(N\bar{g})^{2T}} Z^2 > 0,
\end{aligned}
$$

where $Z > 0$ is the normalizing constant of the model, $Z = \int m_0(dx_0) \prod_{t=1}^{\mathsf{T}} g_t(x_t) f(dx_t|x_{t-1})$. This concludes the proof of Lemma 1.1.

For any fixed T, the bound goes to zero when $N \to \infty$. The proof fails to capture accurately the behaviour of ε in Lemma 1.1 as a function of N and T. Indeed, we observe in the numerical experiments of Section 6 that meeting times decrease when N increases.

2 Proof of Theorem 3.1

The proof is similar to those presented in Rhee (2013), in McLeish (2011), Vihola (2017), and Glynn and Rhee (2014). We can first upper-bound $\mathbb{P}(\tau > n)$, for all $n \geq 2$, using

Lemma 1.1 (e.g. Williams, 1991, exercise E.10.5). We obtain for all $n \geq 2$,

$$\mathbb{P}(\tau > n) \leq (1-\varepsilon)^{n-1}. \tag{7}$$

This ensures that $\mathbb{E}[\tau]$ is finite; and that τ is almost surely finite. We then introduce the random variables $Z_m = \sum_{n=0}^{m} \Delta^{(n)}$ for all $m \geq 1$. Since τ is almost surely finite, and since $\Delta^{(n)} = 0$ for all $n \geq \tau$, then $Z_m \to Z_\tau = H_0$ almost surely when $m \to \infty$. We prove that $(Z_m)_{m\geq 1}$ is a Cauchy sequence in L_2, i.e. $\sup_{m' \geq m} \mathbb{E}\left[(Z_{m'} - Z_m)^2\right]$ goes to 0 as $m \to \infty$. We write

$$\mathbb{E}[(Z_{m'} - Z_m)^2] = \sum_{n=m+1}^{m'} \sum_{\ell=m+1}^{m'} \mathbb{E}[\Delta^{(n)}\Delta^{(\ell)}]. \tag{8}$$

We use Cauchy-Schwarz inequality to write $(\mathbb{E}[\Delta^{(n)}\Delta^{(\ell)}])^2 \leq \mathbb{E}[(\Delta^{(n)})^2]\mathbb{E}[(\Delta^{(\ell)})^2]$, and we note that $(\Delta^{(n)})^2 = \Delta^{(n)}\mathbb{1}(\tau > n)$. Together with Hölder's inequality with $p = 1 + \delta/2$, and $q = (2+\delta)/\delta$, where δ is as in Assumption 3, we can write

$$\mathbb{E}\left[(\Delta^{(n)})^2\right] \leq \mathbb{E}\left[(\Delta^{(n)})^{2+\delta}\right]^{1/(1+\delta/2)} \left((1-\varepsilon)^{\delta/(2+\delta)}\right)^{n-1}.$$

Furthermore, using Assumption 3 and Minkowski's inequality, we obtain the bound

$$\forall n \geq n_0, \qquad \mathbb{E}\left[(\Delta^{(n)})^{2+\delta}\right]^{1/(1+\delta/2)} \leq C_1,$$

where C_1 is independent of n. The above inequalities lead to the terms $\mathbb{E}[\Delta^{(n)}\Delta^{(\ell)}]$ being upper bounded by an expression of the form $C_1\eta^n\eta^\ell$, where $\eta \in (0,1)$. Thus we can compute a bound on Eq. (8), by computing geometric series, and finally conclude that $(Z_m)_{m\geq 1}$ is a Cauchy sequence in L_2.

By uniqueness of the limit, since $(Z_m)_{m\geq 1}$ goes almost surely to H_0, $(Z_m)_{m\geq 1}$ goes to H_0 in L_2. This shows that H_0 has finite first two moments. We can retrieve the expectation of H_0 by

$$\mathbb{E}Z_m = \sum_{n=0}^{m} \mathbb{E}[\Delta^{(n)}] = \mathbb{E}\left[h(X^{(m)})\right] \xrightarrow[m\to\infty]{} \pi(h),$$

according to Assumption 3. This concludes the proof of Theorem 3.1 for H_k with $k = 0$, and a similar reasoning applies for any $k \geq 0$.

References

Chopin, N. and Singh, S. S. (2015). On particle Gibbs sampling. *Bernoulli*, 21(3):1855–1883.

Glynn, P. W. and Rhee, C.-H. (2014). Exact estimation for Markov chain equilibrium expectations. *Journal of Applied Probability*, 51A:377–389.

Lindvall, T. (2002). *Lectures on the coupling method*. Courier Corporation.

McLeish, D. (2011). A general method for debiasing a Monte Carlo estimator. *Monte Carlo methods and applications*, 17(4):301–315.

Rhee, C. (2013). *Unbiased Estimation with Biased Samplers*. PhD thesis, Stanford University.

Vihola, M. (2017). Unbiased estimators and multilevel Monte Carlo. *Operations Research*.

Williams, D. (1991). *Probability with martingales*. Cambridge university press.