Online Appendices Uniform convergence over time of a nested particle filtering scheme for recursive parameter estimation in state–space Markov models

Dan Crisan^{*} and Joaquín Míguez[†] *Department of Mathematics, Imperial College London, d.crisan@imperial.ac.uk. [†]Department of Signal Theory and Communications, Universidad Carlos III de Madrid, joaquin.miguez@uc3m.es.

A Proof of Lemma 1

Before proceeding with the proof of Lemma 1, we need need to introduce some notation, as well as the definition of the Dobrushin contraction coefficient.

For any pair of integers 0 < s < t we can explicitly construct the conditional pdf of the subsequence of observations $y_{s:t}$ given a point $X_s = x_s$ in the state space and a choice parameters $\Theta = \theta$. We denote this density as $g_{s:t,\theta}^{y_{s:t}}(x_s)$, with the notation chosen to make explicit that, for fixed $y_{s:t}$, this is a function of the state value x_s (i.e., it is interpreted as a likelihood). It is not difficult to show that

$$g_{s:t,\theta}^{y_{s:t}}(x_s) = \int \cdots \int \prod_{j=s}^t g_{j,\theta}^{y_j}(x_j) \prod_{l=s+1}^t \tau_{l,\theta}(dx_l | x_{l-1}).$$
(A.1)

We also introduce a specific notation for the conditional distribution of the state X_j conditional on $X_{j-1} = x_{j-1}$, $\Theta = \theta$ and the subsequence of observations from time j up to time t, $y_{j:t}$. For any $j \leq t$, this is a Markov kernel, denoted $k_{j,\theta}^{y_{j:t}}(dx_j|x_{j-1})$, that can be explicitly written as

$$\mathsf{k}_{j,\theta}^{y_{j:t}}(dx_j|x_{j-1}) = \frac{g_{j:t,\theta}^{y_{j:t}}(x_j)\tau_{j,\theta}(dx_j|x_{j-1})}{\int g_{j:t,\theta}^{y_{j:t}}(\tilde{x}_j)\tau_{j,\theta}(d\tilde{x}_j|x_{j-1})}$$
(A.2)

via the Bayes' theorem. If the observation sequence is fixed, then the composite probability measure

$$\mathsf{K}_{s:t+1,\theta}^{y_{s:t}}(dx_{t+1}|x_s) = \int \cdots \int \tau_{t+1,\theta}(dx_{t+1}|x_t) \prod_{j=s+1}^t \mathsf{k}_{j,\theta}^{y_{j:t}}(dx_j|x_{j-1})$$
(A.3)

is a Markov kernel on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

The composite likelihood in (A.1) and the Markov kernel in (A.3) can be used to write integrals w.r.t. the composite map $\Psi_{t+1|s}^{\theta}$ explicitly. To be specific, given a probability measure $\alpha \in \mathcal{P}(\mathcal{X})$, it is an exercise to show that

$$\left(f, \Psi^{\theta}_{t+1|s}(\alpha)\right) = \frac{\left((f, \mathsf{K}^{y_{s:t}}_{s:t+1,\theta})g^{y_{s:t}}_{s:t,\theta}, \alpha\right)}{\left(g^{y_{s:t}}_{s:t,\theta}, \alpha\right)}.$$
(A.4)

The representation in (A.4), together with assumptions A3 and A4, enables the application of standard results from [11] which become instrumental in the analysis of Algorithm 2.

We first define the Dobrushin contraction coefficient [12] for Markov kernels and then show how it can be used to control the difference between between two probability measures $\Psi^{\theta}_{t+1|s}(\alpha)$ and $\Psi^{\theta}_{t+1|s}(\eta)$ which are constructed using the same composite map $\Psi^{\theta}_{t+1|s}$ (and, in particular, the same observation subsequence $y_{s:t+1}$) but different initial conditions $\alpha \neq \eta$.

Definition 5 The Dobrushin contraction coefficient of a Markov kernel K_{θ} from \mathcal{X} onto $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is

$$\beta(K_{\theta}) \triangleq \sup_{x, x' \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})} |K_{\theta}(A|x) - K_{\theta}(A|x')| \le 1.$$

An upper bound for the contraction coefficient of the kernel $\mathsf{K}^{y_{s:t}}_{s:t+1,\theta}$, explicitly given in terms of the constants m, ϵ_{τ} and a in assumptions A4 and A3, is given below.

Lemma 3 If assumptions A3 and A4 hold, then

$$\beta(\mathsf{K}^{y_{s:t}}_{s:t+1,\theta}) \le \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}}\right)^{\lfloor \frac{t-s+1}{m} \rfloor} \tag{A.5}$$

for every $\theta \in D_{\theta}$.

Proof of Lemma 3. Since the inequalities in A3 and A4 are assumed to hold uniformly over the parameter space D_{θ} , the bound in (A.5) follows readily from Proposition 4.3.3 in [11] (see also [11, Corollary 4.3.3]). \Box

Finally, we proceed with the **proof of Lemma 1:** From [11, Proposition 4.3.7] we obtain an upper bound for the difference of integrals that depends on the Dobrushin coefficient of the Markov kernel $K_{s:t+1,\theta}^{y_{s:t}}$, namely

$$\left| (f, \Psi_{t+1|s}^{\theta}(\alpha)) - (f, \Psi_{t+1|s}^{\theta}(\eta)) \right| \le 2 \|f\|_{\infty} \beta(\mathsf{K}_{s:t+1,\theta}^{y_{s:t}}) \left(\sup_{x_s \in \mathcal{X}} \frac{g_{s:t,\theta}^{y_{s:t}}(x_s)}{(g_{s:t,\theta}^{y_{s:t}}, \alpha)} \right) \left| (\tilde{f}_s, \alpha) - (\tilde{f}_s, \eta) \right|,$$
(A.6)

for some $\tilde{f}_s : \mathcal{X} \to \mathbb{R}$ with $\|\tilde{f}_s\| \leq 1$. Moreover, from the definition of the composite likelihood in (A.1) and the assumption $g_{j,\theta}^{y_j} \leq 1$ for every $j \geq 1$ and $\theta \in D_{\theta}$ (in A3), it follows that

$$g_{s:t,\theta}^{y_{s:t}}(x_s) \le (g_{s+m:t,\theta}^{y_{s+m:t}}, \tau_{s+m|s,\theta}(\cdot|x_s))$$
(A.7)

whereas, from the bound $g_{j,\theta}^{y_j}(x) \ge \frac{1}{a}$, for all $j \ge 1$ and $\theta \in D_{\theta}$ (in A3) and the assumption A4, we obtain that

$$(g_{s:t,\theta}^{y_{s:t}}, \alpha) \ge \frac{\epsilon_{\tau}}{a^m} (g_{s+m:t}^{y_{s+m:t}}, \tau_{s+m|s,\theta}(\cdot|\tilde{x}_s))$$
(A.8)

for any $\tilde{x}_s \in \mathcal{X}$. In particular, for $x_s = \tilde{x}_s$, the inequalities (A.7) and (A.8) taken together yield

$$\frac{g_{s:t,\theta}^{y_{s:t}}(x_s)}{(g_{s:t,\theta}^{y_{s:t}},\alpha)} \le \frac{a^m}{\epsilon_\tau}$$

independently of x_s . This, in turn, enables us to rewrite (A.6) as

$$\left| \left(f, \Psi_{t+1|s}^{\theta}(\alpha) \right) - \left(f, \Psi_{t+1|s}^{\theta}(\eta) \right) \right| \le 2 \| f \|_{\infty} \beta(\mathsf{K}_{s:t+1,\theta}^{y_{s:t}}) \frac{a^m}{\epsilon_{\tau}} \left| \left(\tilde{f}_s, \alpha \right) - \left(\tilde{f}_s, \eta \right) \right|.$$
(A.9)

By combining Lemma 3 with (A.9) we readily obtain the inequality (5.3) and complete the proof.

B Proof of Lemma 2

We look into the approximation error $|(f,\xi_{t,\theta_t}^N) - (f,\xi_{t,\theta_t})|$, which can be written as

$$\begin{aligned} \left| (f, \xi_{t,\theta_{t}}^{N}) - (f, \xi_{t,\theta_{t}}) \right| &= \left| \sum_{k=0}^{t-1} \left(f, \Psi_{t|t-k}^{\theta_{t}} \left(\xi_{t-k,\theta_{t-k}}^{N} \right) \right) - \left(f, \Psi_{t|t-k-1}^{\theta_{t}} \left(\xi_{t-k-1,\theta_{t-k-1}}^{N} \right) \right) \right. \\ &+ \left(f, \Psi_{t|0}^{\theta_{t}} \left(\xi_{0,\theta_{0}}^{N} \right) \right) - \left(f, \Psi_{t|0}^{\theta_{t}} \left(\tau_{0} \right) \right) \right| \\ &\leq \sum_{k=0}^{t-1} \left| \left(f, \Psi_{t|t-k}^{\theta_{t}} \left(\xi_{t-k,\theta_{t-k}}^{N} \right) \right) - \left(f, \Psi_{t|t-k-1}^{\theta_{t}} \left(\xi_{t-k-1,\theta_{t-k-1}}^{N} \right) \right) \right| \\ &+ \left| \left(f, \Psi_{t|0}^{\theta_{t}} \left(\xi_{0,\theta_{0}}^{N} \right) \right) - \left(f, \Psi_{t|0}^{\theta_{t}} \left(\tau_{0} \right) \right) \right|, \end{aligned}$$
(B.1)

where the equality follows from a 'telescopic' decomposition of the difference $(f, \xi_{t,\theta_t}^N) - (f, \xi_{t,\theta_t})$. To see this, simply recall that $\xi_{0,\theta_0}^N \equiv \phi_0^N \equiv \tau_0^N$ (independently of θ_0 according to the model in Section 2.2) and note that $\Psi_{t|0}^{\theta_t}(\tau_0) = \xi_{t,\theta_t}$. By way of Minkowski's inequality, (B.1) enables us to express the L_p norm of the approximation error (for $p \ge 1$) as

$$\left\| (f, \xi_{t,\theta_{t}}^{N}) - (f, \xi_{t,\theta_{t}}) \right\|_{p} \leq \sum_{k=0}^{t-1} \left\| \left(f, \Psi_{t|t-k}^{\theta_{t}} \left(\xi_{t-k,\theta_{t-k}}^{N} \right) \right) - \left(f, \Psi_{t|t-k-1}^{\theta_{t}} \left(\xi_{t-k-1,\theta_{t-k-1}}^{N} \right) \right) \right\|_{p} + \left\| \left(f, \Psi_{t|0}^{\theta_{t}} \left(\xi_{0,\theta_{0}}^{N} \right) \right) - \left(f, \Psi_{t|0}^{\theta_{t}} \left(\tau_{0} \right) \right) \right\|_{p},$$
 (B.2)

The last term in the decomposition above can be easily upper bounded using Lemma 1, namely

$$\begin{aligned} \left\| \left(f, \Psi_{t\mid0}^{\theta_t} \left(\xi_{0,\theta_0}^N \right) \right) - \left(f, \Psi_{t\mid0}^{\theta_t} \left(\tau_0 \right) \right) \right\|_p &\leq 2 \|f\|_{\infty} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}} \right)^{\lfloor \frac{t}{m} \rfloor} \frac{a^m}{\epsilon_{\tau}} \left\| \left(\tilde{f}_0, \tau_0^N \right) - \left(\tilde{f}_0, \tau_0 \right) \right\|_p, \\ &\leq 2 \|f\|_{\infty} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}} \right)^{\lfloor \frac{t}{m} \rfloor} \frac{a^m}{\epsilon_{\tau}} \frac{\tilde{C}_0}{\sqrt{N}} \end{aligned} \tag{B.3}$$

where $\|\tilde{f}_0\|_{\infty} \leq 1$ and the second inequality follows readily from the fact that $\tau_0^N = \xi_{0,\theta_0}^N$ is an i.i.d. Monte Carlo approximation of τ_0 (hence, $\tilde{C}_0 < \infty$ is a constant independent of N). For the remaining terms in the sum of (B.2), Lemma 1 yields

$$\left\| \left(f, \Psi_{t|t-k}^{\theta_t} \left(\xi_{t-k,\theta_{t-k}}^N \right) \right) - \left(f, \Psi_{t|t-k-1}^{\theta_t} \left(\xi_{t-k-1,\theta_{t-k-1}}^N \right) \right) \right\|_p \leq 2 \|f\|_{\infty} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}} \right)^{\lfloor \frac{k}{m} \rfloor} \frac{a^m}{\epsilon_{\tau}} \left\| \left(\tilde{f}_{t-k}, \xi_{t-k,\theta_{t-k}}^N \right) - \left(\tilde{f}_{t-k}, \Psi_{t-k}^{\theta_t} \left(\xi_{t-k-1,\theta_{t-k-1}}^N \right) \right) \right\|_p.$$
(B.4)

where $\|\tilde{f}_{t-k}\|_{\infty} \leq 1$.

In order to convert (B.4) into an explicit error rate, we need to derive bounds for errors of the form $\left\| \left(h, \xi_{t-k,\theta_{t-k}}^N\right) - \left(h, \Psi_{t-k}^{\theta_t}\left(\xi_{t-k-1,\theta_{t-k-1}}^N\right)\right) \right\|_p$, where $h : \mathcal{X} \to \mathbb{R}$ with $\|h\|_{\infty} \leq 1$. With this aim, we consider the triangular inequality

$$\left\| \left(h, \xi_{t-k,\theta_{t-k}}^{N}\right) - \left(h, \Psi_{t-k}^{\theta_{t}}\left(\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)\right) \right\|_{p} \leq \left\| \left(h, \xi_{t-k,\theta_{t-k}}^{N}\right) - E\left[\left(h, \xi_{t-k,\theta_{t-k}}^{N}\right) |\mathcal{G}_{t-k}\right] \right\|_{p} + \left\| E\left[\left(h, \xi_{t-k,\theta_{t-k}}^{N}\right) |\mathcal{G}_{t-k}\right] - \left(h, \Psi_{t-k}^{\theta_{t}}\left(\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)\right) \right\|_{p}, \quad (B.5)$$

where $\mathcal{G}_{t-k} = \sigma\left(x_{0:t-k-1}^{(n)}, \bar{x}_{1:t-k-1}^{(n)}, \{\theta_s\}_{s \ge 0}; 1 \le n \le N\right)$ is the σ -algebra generated by the random variables between brackets, and analyse the two terms on the right hand side separately.

For the first term on the right hand side of (B.5), we note that

$$\left(h, \xi_{t-k,\theta_{t-k}}^{N}\right) - E\left[\left(h, \xi_{t-k,\theta_{t-k}}^{N}\right) | \mathcal{G}_{t-k}\right] = \frac{1}{N} \sum_{n=1}^{N} \bar{S}_{t-k}^{(n)},$$

where

$$\bar{S}_{t-k}^{(n)} = h(\bar{x}_{t-k}^{(n)}) - E\left[h(\bar{x}_{t-k}^{(n)})|\mathcal{G}_{t-k}\right], \quad n = 1, ..., N,$$

are zero-mean and conditionally (on \mathcal{G}_{t-k}) independent r.v.'s. Therefore it is straightforward to show that

$$E\left[\left|\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)-E\left[\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)|\mathcal{G}_{t-k}\right]\right|^{p}|\mathcal{G}_{t-k}\right]=E\left[\left|\frac{1}{N}\sum_{n=1}^{N}\bar{S}_{t-k}^{(n)}\right|^{p}|\mathcal{G}_{t-k}\right]\leq\frac{c^{p}}{N^{\frac{p}{2}}}\tag{B.6}$$

for some constant c > 0 independent of N and independent of the distribution of the variables $\bar{S}_{t-k}^{(n)}$, n = 1, ..., N (in particular, independent of the sequence $\{\theta_t\}_{t\geq 0}$). Taking expectations on both sides of (B.6), and then exponentiating by $\frac{1}{p}$, yields

$$\left\| \left(h, \xi_{t-k,\theta_{t-k}}^N \right) - E\left[\left(h, \xi_{t-k,\theta_{t-k}}^N \right) | \mathcal{G}_{t-k} \right] \right\|_p \le \frac{c}{\sqrt{N}}.$$
(B.7)

To find a rate for the second term in (B.5), we note that

$$E\left[\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)|\mathcal{G}_{t-k}\right] = \frac{\left(g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}}\left(h,\tau_{t-k,\theta_{t-k}}\right),\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)}{\left(g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}},\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)}$$
(B.8)

whereas

$$\left(h, \Psi_{t-k}^{\theta_{t}}\left(\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)\right) = \frac{\left(g_{t-k-1,\theta_{t}}^{y_{t-k-1}}\left(h, \tau_{t-k,\theta_{t}}\right), \xi_{t-k-1,\theta_{t-k-1}}^{N}\right)}{\left(g_{t-k-1,\theta_{t}}^{y_{t-k-1}}, \xi_{t-k-1,\theta_{t-k-1}}^{N}\right)}.$$
(B.9)

Subtracting (B.9) from (B.8) and then rearranging terms yields

$$E\left[\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)|\mathcal{G}_{t-k}\right] - \left(h,\Psi_{t-k}^{\theta_{t}}\left(\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)\right) = \frac{\left(g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}}\left(h,\tau_{t-k,\theta_{t-k}}\right) - g_{t-k-1,\theta_{t}}^{y_{t-k-1}}\left(h,\tau_{t-k,\theta_{t}}\right),\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)}{\left(g_{t-k-1,\theta_{t}}^{y_{t-k-1}},\xi_{t-k-1,\theta_{t-k-1}}^{k}\right)} + \frac{E\left[\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)|\mathcal{G}_{t-k}\right] \times \left(g_{t-k-1,\theta_{t}}^{y_{t-k-1}} - g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}},\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)}{\left(g_{t-k-1,\theta_{t}}^{y_{t-k-1}},\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)},$$

hence

$$\left| E \left[\left(h, \xi_{t-k,\theta_{t-k}}^{N} \right) | \mathcal{G}_{t-k} \right] - \left(h, \Psi_{t-k}^{\theta_{t}} \left(\xi_{t-k-1,\theta_{t-k-1}}^{N} \right) \right) \right| \leq a \times \left(\left| g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}} \left(h, \tau_{t-k,\theta_{t-k}} \right) - g_{t-k-1,\theta_{t}}^{y_{t-k-1}} \left(h, \tau_{t-k,\theta_{t}} \right) \right|, \xi_{t-k-1,\theta_{t-k-1}}^{N} \right) + a \times \left(\left| g_{t-k-1,\theta_{t}}^{y_{t-k-1}} - g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}} \right|, \xi_{t-k-1,\theta_{t-k-1}}^{N} \right), \tag{B.10}$$

where we have used the obvious bounds $E\left[\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)|\mathcal{G}_{t-k}\right] \leq \|h\|_{\infty} \leq 1$ and, from assumption A3, $\left(g_{t-k-1,\theta_{t}}^{y_{t-k-1}},\xi_{t-k-1,\theta_{t-k-1}}^{N}\right) \geq a^{-1}.$

From assumption A5, the likelihoods $g_{t,\theta}^{y_t}(x)$ are Lipschitz in the parameter θ , with constant L_g independent of t and x. In particular,

$$\sup_{x \in \mathcal{X}, t \ge T} \left| g_{t-k-1,\theta_t}^{y_{t-k-1}}(x) - g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}}(x) \right| \le L_g \|\theta_t - \theta_{t-k-1}\|.$$
(B.11)

Also from assumption A5, the kernels $\tau_{t,\theta}(dx|x) \in \mathcal{P}(\mathcal{X})$ are endowed with densities w.r.t. the Lebesgue measure, hence we can write

$$\left| g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}}(x) \left(h,\tau_{t-k,\theta_{t-k}}\right)(x) - g_{t-k-1,\theta_{t}}^{y_{t-k-1}}(x) \left(h,\tau_{t-k,\theta_{t}}\right)(x) \right| = \\ \left| g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}}(x) \int h(x') \tau_{t-k,\theta_{t-k}}^{x}(x') dx' - g_{t-k-1,\theta_{t}}^{y_{t-k-1}}(x) \int h(x') \tau_{t-k,\theta_{t}}^{x}(x') dx' \right|$$

and a simple triangle inequality yields

$$\begin{vmatrix} g_{t-k-1}^{y_{t-k-1}}(x)\left(h,\tau_{t-k,\theta_{t-k}}\right)(x) - g_{t-k-1,\theta_{t}}^{y_{t-k-1}}(x)\left(h,\tau_{t-k,\theta_{t}}\right)(x) \end{vmatrix} \leq \\ \left| \left(g_{t-k-1,\theta_{t-k-1}}^{y_{t-k-1}}(x) - g_{t-k-1,\theta_{t-k}}^{y_{t-k-1}}(x) \right) \int h(x')\tau_{t-k,\theta_{t-k}}^{x}(x')dx' \right| + \\ \left| \int h(x')\left(g_{t-k-1,\theta_{t-k}}^{y_{t-k-1}}(x)\tau_{t-k,\theta_{t-k}}^{x}(x') - g_{t-k-1,\theta_{t}}^{y_{t-k-1}}(x)\tau_{t-k,\theta_{t}}^{x}(x') \right) dx' \right| \leq \\ \left| \left(L_{g} \lor L_{g,\tau} \right) \left(\left\| \theta_{t-k-1} - \theta_{t-k} \right\| + \left\| \theta_{t} - \theta_{t-k} \right\| \right) \right), \end{aligned} \right|$$
(B.12)

where the second inequality is satisfied because the product $g_{t,\theta}^{y_t} \tau_{t,\theta'}^x(x')$ is Lipschitz in θ for every $t \ge 1$ and $x, x' \in \mathcal{X}$ (a consequence of assumption A5) with constant $L_{g,\tau}$.

If we substitute (B.11) and (B.12) back into (B.10) we obtain

$$\left| E\left[\left(h, \xi_{t-k,\theta_{t-k}}^N \right) | \mathcal{G}_{t-k} \right] - \left(h, \Psi_{t-k}^{\theta_t} \left(\xi_{t-k-1,\theta_{t-k-1}}^N \right) \right) \right| \le 2aL \sum_{j=0}^k \|\theta_{t-j} - \theta_{t-j-1}\|$$
(B.13)

where we have introduced the constant $L = \max\{L_g, L_{g,\tau}\}$ and taken advantage of the straightforward inequality $\|\theta_t - \theta_{t-k-1}\| \leq \sum_{j=0}^k \|\theta_{t-j} - \theta_{t-j-1}\|$. Raising both sides of (B.13) to power p and then taking

expectations yields

$$E\left[\left|E\left[\left(h,\xi_{t-k,\theta_{t-k}}^{N}\right)|\mathcal{G}_{t-k}\right] - \left(h,\Psi_{t-k}^{\theta_{t}}\left(\xi_{t-k-1,\theta_{t-k-1}}^{N}\right)\right)\right|^{p}\right] \leq (2aL)^{p}E\left[\left|\sum_{j=0}^{k} \|\theta_{t-j} - \theta_{t-j-1}\|\right|^{p}\right] \leq (2aL(k+1))^{p} \times \frac{1}{k+1}\sum_{j=0}^{k}E\left[\|\theta_{t-j} - \theta_{t-j-1}\|^{p}\right],$$
(B.14)

where (B.14) follows from Jensen's inequality. Combining (B.14) with Proposition 1 we arrive at

$$\left\| E\left[\left(h, \xi_{t-k,\theta_{t-k}}^N\right) | \mathcal{G}_{t-k} \right] - \left(h, \Psi_{t-k}^{\theta_t} \left(\xi_{t-k-1,\theta_{t-k-1}}^N \right) \right) \right\|_p \le 2aL(k+1)\frac{c_\kappa}{\sqrt{N}}, \tag{B.15}$$

where $c_{\kappa} < \infty$ is a constant independent of N, t and $\{\theta_n\}_{n \ge 0}$.

If we now insert (B.7) and (B.15) into (B.5) we obtain the relationship

$$\left\| \left(h, \xi_{t-k,\theta_{t-k}}^N\right) - \left(h, \Psi_{t-k}^{\theta_t}\left(\xi_{t-k-1,\theta_{t-k-1}}^N\right)\right) \right\|_p \le \frac{c+2aL(k+1)c_\kappa}{\sqrt{N}},\tag{B.16}$$

where the numerator is finite and constant w.r.t. N, $\{\theta_n\}_{n\geq 0}$ and t. At this point, we only need to substitute the latter inequality backwards. Indeed, if we plug (B.16), with $h = \tilde{f}_{t-k}$, into (B.4) and then substitute the resulting bound, together with (B.3), into (B.2), we arrive at

$$\left\| (f,\xi_{t,\theta_{t}}^{N}) - (f,\xi_{t,\theta_{t}}) \right\|_{p} \leq \frac{2\|f\|_{\infty} a^{m} \epsilon_{\tau}^{-1}}{\sqrt{N}} \sum_{k=0}^{t} \left(1 - \frac{\epsilon_{\tau}^{2}}{a^{m-1}} \right)^{\lfloor \frac{k}{m} \rfloor} (\bar{C}_{0} + \bar{C}_{1}k), \tag{B.17}$$

where $\bar{C}_0 = c + 2aLc_{\kappa}$ and $\bar{C}_1 = \tilde{C}_0 \vee 2aLc_{\kappa}$.

What remains to be proved is that the sum in (B.17) admits an upper bound $\bar{C} < \infty$ independent of t. To show this, we decompose

$$\sum_{k=0}^{t} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}}\right)^{\lfloor \frac{k}{m} \rfloor} (\bar{C}_0 + \bar{C}_1 k) = \bar{C}_0 \sum_{k=0}^{t} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}}\right)^{\lfloor \frac{k}{m} \rfloor} + \bar{C}_1 \sum_{k=0}^{t} k \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}}\right)^{\lfloor \frac{k}{m} \rfloor}$$
(B.18)

and note that each term in (B.18) can be written as a sum of convergent series. Indeed, for the first term we have

$$\sum_{k=0}^{t} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}}\right)^{\lfloor \frac{k}{m} \rfloor} \leq m \sum_{k=0}^{\infty} \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}}\right)^k$$
(B.19)

$$= ma^{m-1}\epsilon_{\tau}^{-2}, \tag{B.20}$$

where the inequality (B.19) is obtained from the identity $\sum_{k=0}^{\infty} r^{\lfloor \frac{k}{m} \rfloor} = m \sum_{k=0}^{\infty} r^k$ (for any $r \in (0,1)$)

and (B.20) follows from the limit of the geometric series. For the second term in (B.18) we have

$$\sum_{k=0}^{t} k \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}} \right)^{\lfloor \frac{k}{m} \rfloor} \leq 2m \sum_{k=0}^{\infty} \left\lfloor \frac{k}{m} \right\rfloor \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}} \right)^{\lfloor \frac{k}{m} \rfloor}$$
(B.21)

$$= 2m^2 \sum_{k=0}^{\infty} k \left(1 - \frac{\epsilon_{\tau}^2}{a^{m-1}} \right)^k, \qquad (B.22)$$

$$= 2m^2 \frac{1 - \epsilon_\tau^2 a^{-(m-1)}}{\epsilon_\tau^2 a^{-2(m-1)}}, \tag{B.23}$$

where (B.21) follows from the inequality $k \leq 2m \lfloor \frac{k}{m} \rfloor$ (for k = 0, 1, 2, ... and $m \geq 1$), (B.22) holds because of the identity $\sum_{k=0}^{\infty} \lfloor \frac{k}{m} \rfloor r^{\lfloor \frac{k}{m} \rfloor} = m \sum_{k=0}^{\infty} kr^k$ (for any $r \in (0, 1)$) and (B.23) is readily obtained from the limit $\sum_{k=0}^{\infty} kr^k = \frac{r}{(1-r)^2}$ (for |r| < 1).

To conclude the proof, we simply put (B.17), (B.18), (B.20) and (B.23) together, to obtain the desired inequality (5.5) with

$$\bar{C} = 2\|f\|_{\infty} a^m \epsilon_{\tau}^{-1} \left(\bar{C}_0 m a^{m-1} \epsilon_{\tau}^{-2} + 2\bar{C}_1 m^2 \frac{1 - \epsilon_{\tau}^2 a^{-(m-1)}}{\epsilon_{\tau}^2 a^{-2(m-1)}} \right) \le 4\|f\|_{\infty} (\bar{C}_0 \vee \bar{C}_1) \epsilon_{\tau}^{-3} a^{3m} e^{-2\pi i \epsilon_{\tau}^2} e^{-2\pi i \epsilon_{\tau}^2}$$

and $\bar{C}_0 \vee \bar{C}_1 \leq a(c + \tilde{C}_0 + 2Lc_\kappa).$

C A proof for inequality (5.27)

We need to prove that $||(v, \bar{\mu}_{n-1}^N) - (v, \mu_{n-1}^N)||_p \leq \frac{s_1 ||v||_{\infty}}{\sqrt{N}}$ for some $s_1 < \infty$ independent of N and $v \in B(D_{\theta})$.

Recall that we draw the particles $\bar{\theta}_n^{(i)}$, i = 1, ..., N, independently from the kernels $\kappa_{N,p}^{\theta_{n-1}^{(i)}}$, i = 1, ..., N, respectively, and start from the triangle inequality

$$\|(v,\bar{\mu}_{n-1}^N) - (v,\mu_{n-1}^N)\|_p \le \|(v,\bar{\mu}_{n-1}^N) - (v,\kappa_{N,p}\mu_{n-1}^N)\|_p + \|(v,\kappa_{N,p}\mu_{n-1}^N) - (v,\mu_{n-1}^N)\|_p$$
(C.1)

where

$$(v, \kappa_{N, \mathsf{p}} \mu_{n-1}^N) = \frac{1}{N} \sum_{i=1}^N \int v(\theta) \kappa_{N, \mathsf{p}}^{\theta_{n-1}^{(i)}}(d\theta),$$

and then analyse the two terms on the right hand side of (C.1) separately.

Let \mathcal{G}_{n-1} be the σ -algebra generated by the random particles $\{\bar{\theta}_{1:n-1}^{(i)}, \theta_{0:n-1}^{(i)}\}_{1 \leq i \leq N}$. Then

$$E\left[(v,\bar{\mu}_{n-1}^{N})|\mathcal{G}_{n-1}\right] = \frac{1}{N}\sum_{i=1}^{N}\int v(\theta)\kappa_{N,\mathsf{p}}^{\theta_{n-1}^{(i)}}(d\theta) = (v,\kappa_{N,\mathsf{p}}\mu_{n-1}^{N})$$

and the difference $(v, \bar{\mu}_{n-1}^N) - (v, \kappa_{N,p} \mu_{n-1}^N)$ can be written as

$$(v, \bar{\mu}_{n-1}^N) - (v, \kappa_{N, \mathsf{p}} \mu_{n-1}^N) = \frac{1}{N} \sum_{i=1}^N \bar{Z}_{n-1}^{(i)},$$

where the random variables $\bar{Z}_{n-1}^{(i)} = v(\bar{\theta}_n^{(i)}) - E[v(\bar{\theta}_n^{(i)})|\mathcal{G}_{n-1}], i = 1, ..., N$, are conditionally independent (given \mathcal{G}_{n-1}), have zero mean and can be bounded as $|\bar{Z}_{n-1}^{(i)}| \leq 2||v||_{\infty}$. As a consequence, it is an exercise in combinatorics to show that

$$E\left[\left|(v,\bar{\mu}_{n-1}^{N}) - (v,\kappa_{N,\mathsf{p}}\mu_{n-1}^{N})\right|^{p}|\mathcal{G}_{n-1}\right] = E\left[\left|\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_{n-1}^{(i)}\right|^{p}|\mathcal{G}_{n-1}\right] \le \frac{\tilde{c}_{1}^{p}||v||_{\infty}^{p}}{N^{\frac{p}{2}}},\tag{C.2}$$

where \tilde{c}_1 is a constant independent of N, n and v (actually, independent of the distribution of the $\bar{Z}_{n-1}^{(i)}$'s). From (C.2) we readily obtain that

$$\|(v,\bar{\mu}_{n-1}^{N}) - (v,\kappa_{N,p}\mu_{n-1}^{N})\|_{p} \le \frac{\tilde{c}_{1}\|v\|_{\infty}}{\sqrt{N}}.$$
(C.3)

For the remaining term in (C.1), namely, $\|(v, \kappa_{N,p}\mu_{n-1}^N) - (v, \mu_{n-1}^N)\|_p$, we simply note that

$$\begin{aligned} \left| (v, \kappa_{N, \mathbf{p}} \mu_{n-1}^{N}) - (v, \mu_{n-1}^{N}) \right| &= \left| \frac{1}{N} \sum_{i=1}^{N} \int \left(v(\theta) - v(\theta_{n-1}^{(i)}) \right) \kappa_{N, \mathbf{p}}^{\theta_{n-1}^{(i)}}(d\theta) \right| \\ &\leq \frac{1}{N} \sum_{i=1}^{N} \int \left| v(\theta) - v(\theta_{n-1}^{(i)}) \right| \kappa_{N, \mathbf{p}}^{\theta_{n-1}^{(i)}}(d\theta) \leq \frac{2 \|v\|_{\infty}}{\sqrt{N}}, \end{aligned}$$
(C.4)

where the last inequality follows from Proposition 1.

Substituting the inequalities (C.3) and (C.4) into Eq. (C.1) yields the desired conclusion, viz., Eq. (5.27), with constant $s_1 = 2 + \tilde{c}_1$ independent of N.

D Additional results on the numerical experiment of Figure 1

In order to provide additional numerical evidence of the stability of the posterior-mean estimates of the parameters over time, Figure 5 shows the normalised posterior standard deviation (NSTD) of the parameter estimates for the same simulation run as in Figure 1. At each time n, this is computed for the j-th parameter, j = 1, ..., 4, as

$$NSTD_{j,n} = \frac{\sqrt{\sum_{i=1}^{N} w_n^{(i)} (\bar{\theta}_{j,n}^{(i)} - \hat{\theta}_{j,n}^N)^2}}{\theta_j^*},$$

where θ_j^* is the true value of the *j*-th parameter (namely, $\theta_1^* = S = 10, \theta_2^* = R = 28, \theta_3^* = B = \frac{8}{3}$ and $\theta_4^* = k_o = 0.8$). Again, the NSTD is a random statistic and it displays fluctuations, however it can be seen that their amplitudes remain bounded and there is no apparent increase over time.



Figure 5: Evolution of the normalised posterior standard deviation of the Lorenz 63 model parameters S, R, B and k_o over time. The horizontal axes are labeled with continuous time units. After Euler's discretisation, each continuous time unit amounts to 1,000 discrete time steps, with one observation vector every 40 discrete-time steps. The number of particles is N = M = 300.