Gerber, M., & Chopin, N. (2017). Convergence of sequential quasi-Monte Carlo smoothing algorithms. *Bernoulli*, *23*(4B), 2951-2987. https://doi.org/10.3150/16-BEJ834

Peer reviewed version

Link to published version (if available):
10.3150/16-BEJ834

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Bernoulli Society at http://projecteuclid.org/euclid.bj/1495505081. Please refer to any applicable terms of use of the publisher.

Convergence of sequential Quasi-Monte Carlo smoothing algorithms

Mathieu Gerber* Nicolas Chopin†

[17] recently introduced Sequential quasi-Monte Carlo (SQMC) algorithms as an efficient way to perform filtering in state-space models. The basic idea is to replace random variables with low-discrepancy point sets, so as to obtain faster convergence than with standard particle filtering. [17] describe briefly several ways to extend SQMC to smoothing, but do not provide supporting theory for this extension. We discuss more thoroughly how smoothing may be performed within SQMC, and derive convergence results for the so-obtained smoothing algorithms. We consider in particular SQMC equivalents of forward smoothing and forward filtering backward sampling, which are the most well-known smoothing techniques. As a preliminary step, we provide a generalization of the classical result of [22] on the transformation of QMC point sets into low discrepancy point sets with respect to non uniform distributions. As a corollary of the latter, we note that we can slightly weaken the assumptions to prove the consistency of SQMC.

Keywords: Hidden Markov models; Low discrepancy; Particle filtering; Quasi-Monte Carlo; Sequential quasi-Monte Carlo; Smoothing; State-space models.

1. Introduction

State-space models are popular tools to model real life phenomena in many fields such as Economics, Engineering and Neuroscience. These models are mainly used for extracting information about a hidden Markov process $(\mathbf{x}_t)_{t\geq 0}$ of interest from a set of observations $\mathbf{y}_{0:T} := (\mathbf{y}_0, \dots, \mathbf{y}_T)$. In practice, this typically translates to the estimation of $p(\mathbf{x}_t|\mathbf{y}_{0:t})$, the distribution of $\mathbf{x}_t$ given the data $\mathbf{y}_{0:t}$, $0 \leq t \leq T$ (called the *filtering* distribution), and/or to $p(\mathbf{x}_{0:T}|\mathbf{y}_{0:T})$ (called the *smoothing* distribution). However, these distributions

*Harvard University, Department of Statistics, Science Center 7th floor, One Oxford Street, Cambridge, MA 02138-2901, USA mathieugerber@fas.harvard.edu

†CREST-ENSAE, 3 Avenue Pierre Larousse, 92245 Malakoff, FRANCE, nicolas.chopin@ensae.fr (corresponding author)

are intractable in most cases, and require to be approximated in some way, the most popular being particle filtering (Sequential Monte Carlo). See e.g. the books of [13], [6] for more background on state-space models and particle filters.

Recently, [17] introduced sequential quasi-Monte Carlo (SQMC) as an efficient alternative to particle filtering. (SQMC is related to the array-RQMC algorithm, [24, 15, 25, 23], as discussed in the body of the paper.) Essentially, SQMC amounts to replacing the random variates generated at each iteration of a particle filter with a QMC (low-discrepancy) point set; that is a set of N points that are selected so as to cover more evenly the space that random variates would; see e.g. the books of [26], [27] for more background on QMC. [17] established that, for some constructions of RQMC (randomised QMC) point sets, the convergence rate of SQMC (with respect to N, the number of simulations) is at worst $\mathcal{O}_P(N^{-1/2})$, while it is $o_P(N^{-1/2})$ on the class of continuous and bounded functions. (This of course compares favourably to the $\mathcal{O}_P(N^{-1/2})$ rate of particle filtering.) In addition, the numerical results of [17] show that SQMC dramatically outperforms particle filtering in several applications.

One important question that remains however is how to use SQMC to obtain smoothing estimates that converge as $N \to +\infty$. Smoothing is recognised as a more difficult problem than filtering [4]. Smoothing algorithms typically require extra steps on top of particle filtering (such as a backward pass), and often cost $\mathcal{O}(N^2)$ (but some variants cost $\mathcal{O}(N)$, as discussed later).

This paper discusses existing smoothing algorithms, explains how they may be adapted to SQMC, and presents convergence results for the corresponding SQMC smoothing algorithms. We first study forward smoothing, where trajectories are carried forward in the particle filter, and show that this approach leads to consistent estimates in SQMC. Then, we derive a SQMC version of forward filtering backward sampling (where complete trajectories are simulated from the positions simulated by a particle filter, see 18), and establish convergence results for the so obtained smoothing estimates. We also consider the marginal version of backward sampling, which usually allows for a more precise estimation of marginal smoothing distributions.

The rest of this paper is organized as follows. Section 2 introduces the model and the notations considered in this work, and give a short description of SQMC. Section 3 contains some preliminary results that will be needed to study SQMC smoothing. We first present a new consistency result for the forward step, which has the advantage to rely on weaker assumptions than in [17], and state a result relative to the backward decomposition and SQMC estimation of the smoothing distribution. Then, we provide a generalization of the classical result of [22] on the transformation of QMC point sets into low discrepancy point sets with respect to non uniform distributions that is essential to the analysis of QMC smoothing algorithms. This section ends with some results on the conversion of discrepancies through the Hilbert space filling curve. In Section 4 we establish the consistency of QMC forward smoothing while our results on QMC forward-backward smoothing are given Section 5. In Section 6 a numerical study examines the performance of the QMC smoothing strategies discussed in this work while Section 7 concludes.

2. Preliminaries

2.1. Feynman-Kac formalism

Let $(\mathbf{x}_t)_{t\geq 0}$ be a Markov chain, defined on a space $\mathcal{X} \subseteq \mathbb{R}^d$ (equipped with Lebesgue measure), with initial distribution $m_0(\mathrm{d}\mathbf{x}_0)$, transition kernel $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$, $t \geq 1$, and let $(G_t)_{t\geq 0}$ a sequence of (measurable) potential functions, $G_0 : \mathcal{X} \to \mathbb{R}^+$, $G_t : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$. As in [17], and most of the QMC literature, we take $\mathcal{X} = [0,1)^d$, but see Section 3 of [17] for how to generalise our results to unbounded state spaces.

For this Feynman-Kac model $(m_t, G_t)_{t\geq 0}$, let $\overline{\mathbb{Q}}_t$ and $\mathbb{Q}_t$ be the probability measures on $\mathcal{X}$ such that, for any bounded measurable function $\varphi : \mathcal{X} \to \mathbb{R}$,

$$\overline{\mathbb{Q}}_t(\varphi) = \frac{1}{Z_{t-1}} \mathbb{E}\left[\varphi(\mathbf{x}_t) G_0(\mathbf{x}_0) \prod_{s=1}^{t-1} G_s(\mathbf{x}_{s-1}, \mathbf{x}_s)\right]$$

$$\mathbb{Q}_t(\varphi) = \frac{1}{Z_t} \mathbb{E}\left[\varphi(\mathbf{x}_t) G_0(\mathbf{x}_0) \prod_{s=1}^{t} G_s(\mathbf{x}_{s-1}, \mathbf{x}_s)\right]$$

$$Z_t = \mathbb{E}\left[G_0(\mathbf{x}_0) \prod_{s=1}^{t} G_s(\mathbf{x}_{s-1}, \mathbf{x}_s)\right]$$

where expectations are with respect to the law of Markov chain $(\mathbf{x}_t)$, and empty products equal one. Similarly, let $\widetilde{\mathbb{Q}}_t$ be the probability measure on $\mathcal{X}^{t+1}$ such that, for any bounded test function $\varphi : \mathcal{X}^{t+1} \to \mathbb{R}$,

$$\widetilde{\mathbb{Q}}_t(\varphi) = \frac{1}{Z_t} \mathbb{E}\left[\varphi(\mathbf{x}_{0:t}) G_0(\mathbf{x}_0) \prod_{s=1}^{t} G_s(\mathbf{x}_{s-1}, \mathbf{x}_s)\right].$$

In the sequel, the notation $0:t$ is used to denote the set of integers $\{0, \ldots, t\}$ and $\mathbf{x}_{0:t}$ denotes the collection $\{\mathbf{x}_s\}_{s=0}^t$. Similarly, in what follows we use the shorthand $\mathbf{x}^{1:N}$ for a collection $\{\mathbf{x}^n\}_{n=1}^N$ of N points in $\mathbb{R}^d$, and $\mathbf{x}_{0:t}^{1:N}$ for collection $\{\mathbf{x}_{0:t}^n\}_{n=1}^N$ of N points in $\mathbb{R}^{(t+1)d}$. Finally, for a probability measure $\pi \in \mathcal{P}(\mathcal{X})$, with $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$ absolutely continuous with respect to the Lebesgue measure, $\pi(\varphi)$ denotes the expectation of $\varphi(\mathbf{x})$ under π.

2.2. Connection with state-space models and smoothing

The Feynman-Kac formalism of the previous section (see the book of [9] for a more in-depth presentation) is rather abstract, but it has the advantage of representing in a very generic manner the recursive quantities computed by most types of particle algorithms, in particular those aimed at the sequential analysis of state-space models.

Concretely, consider a state-space model consisting of an unobserved Markov chain $(\mathbf{x}_t)$, with initial distribution $p_0(\mathrm{d}\mathbf{x}_0)$ and Markov kernel $p_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$, and an observed process $(\mathbf{y}_t)$, such that $\mathbf{y}_t$, conditional on $\mathbf{x}_t$, is independent of the other variables, $\mathbf{y}_s$ and $\mathbf{x}_s$ for $s \neq t$, and has conditional density (with respect to an appropriate measure, e.g.

Lebesgue) $f^Y(\mathbf{y}_t|\mathbf{x}_t)$. For instance, in tracking and navigation applications, $\mathbf{x}_t$ would represent the position of a moving object (e.g. a ship), and $\mathbf{y}_t$ a noisy measurement of this position (e.g. by a radar); see e.g. the books of [13] and [6] for other examples of state-space models.

By taking

$$G_0(\mathbf{x}_0) = f^Y(\mathbf{y}_0|\mathbf{x}_0)\frac{m_0(\mathrm{d}\mathbf{x}_0)}{p_t(\mathrm{d}\mathbf{x}_0)}, \quad G_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = f^Y(\mathbf{y}_t|\mathbf{x}_t)\frac{m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)}{p_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)},$$

where the second factor in both cases is a Radon-Nikodym derivative (of $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ relative to $p_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ in the latter case, assuming of course, this derivative is well defined), we obtain as $\mathbb{Q}_t(\mathrm{d}\mathbf{x}_t)$ the filtering distribution of the model, that is, the law of $\mathbf{x}_t$ conditional on $\mathbf{y}_{0:t}$. Similarly $\overline{\mathbb{Q}}_t(\mathrm{d}\mathbf{x}_t)$ is the predictive distribution (the law of $\mathbf{x}_t|\mathbf{y}_{0:t-1}$), and $\widetilde{\mathbb{Q}}_t(\mathrm{d}\mathbf{x}_{0:t})$ is the object of interest in this work, namely the smoothing distribution (the law of $\mathbf{x}_{0:t}|\mathbf{y}_{0:t}$). In addition, Z_t is the marginal likelihood of observations $\mathbf{y}_{0:t}$.

A particle algorithm corresponding to a given Feynman-Kac model simulates particles according to the Markov kernel $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$. One may take $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t) = p_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$, in which case G_t simplifies to $G_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = f^Y(\mathbf{y}_t|\mathbf{x}_t)$; i.e. G_t depends only on $\mathbf{x}_t$. The corresponding algorithm is usually called the bootstrap filter, after [19]. However, one may often construct more efficient particle algorithms by taking a $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ that differs from $p_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ [14].

2.3. Extreme norm and QMC point sets

As in [17], our consistency results are stated in term of the *extreme* norm, defined, for two probability measures π_1 and π_2 on $[0,1)^d$, by

$$\|\pi_1 - \pi_2\|_{\mathrm{E}} = \sup_{B \in \mathcal{B}_{[0,1)^d}} |\pi_1(B) - \pi_2(B)|$$

where

$$\mathcal{B}_{[0,1)^d} = \{B = \prod_{i=1}^d [a_i, b_i], \ 0 \le a_i < b_i < 1\}.$$

Note that $\|\pi_N - \pi\|_{\mathrm{E}} \to 0$ implies that $\pi_N(\varphi) \to \pi(\varphi)$ for any bounded and continuous function φ (by portmanteau lemma, see e.g. Lemma 2.2, p.6 of [36]). In words, consistency for the extreme norm implies consistency of estimates for test functions φ that are bounded and continuous.

The extreme norm is natural in QMC contexts since it can be viewed as the generalization of the *extreme discrepancy* of a point set $\mathbf{u}^{1:N}$ in $[0,1)^d$, defined by

$$D(\mathbf{u}^{1:N}) = \|\mathcal{S}(\mathbf{u}^{1:N}) - \lambda_d\|_{\mathrm{E}}$$

where λ_d denotes the Lebesgue measure on $\mathbb{R}^d$ and $\mathcal{S}$ is the operator

$$\mathcal{S} : \mathbf{u}^{1:N} \to \frac{1}{N}\sum_{n=1}^N \delta_{\mathbf{u}^n}.$$

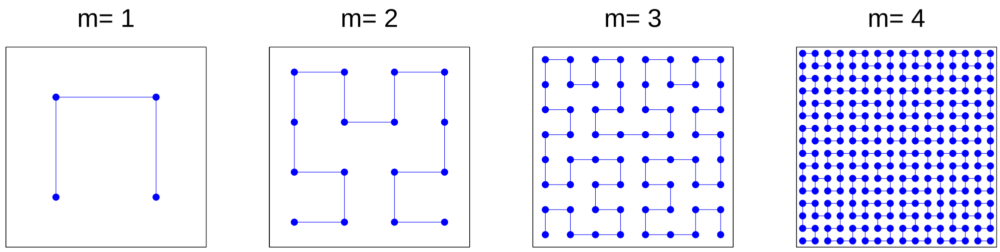

Figure 1: First four iterates of sequence H_m, the limit of which is the Hilbert curve H, for $d = 2$ (source: [21])

The extreme discrepancy therefore measures how a point set spreads evenly over $[0,1)^d$ and is used to define formally QMC point sets. To be more specific, $\mathbf{u}^{1:N}$ is a QMC point set in $[0,1)^d$ if $D(\mathbf{u}^{1:N}) = \mathcal{O}(N^{-1}(\log N)^d)$. Note that, for a sample $\mathbf{u}^{1:N}$ of N IID uniform random numbers in $[0,1)^d$, $D(\mathbf{u}^{1:N}) = \mathcal{O}(N^{-1/2}\log\log N)$ almost surely by the law of iterated logarithm [see e.g. 28, page 167]. There exist many constructions of QMC point sets in the literature [see 28, 11, for more details on this topic] and, although we write $\mathbf{u}^{1:N}$ rather than $\mathbf{u}^{N,1:N}$, $\mathbf{u}^{1:N}$ may not necessarily be the N first points of a fixed sequence, i.e. one may have $\mathbf{u}^{N,N-1} \neq \mathbf{u}^{N-1,N-1}$. However, it is worth keeping in mind that all the results presented in this paper hold both for point sets and sequences.

Even if in this work we are mainly interested in consistency results (which hold for deterministic point sets $\mathbf{u}^{1:N}$), we will sometimes refer to randomized QMC (RQMC) point sets. Formally, $\mathbf{u}^{1:N}$ is RQMC point set if it is a QMC point set with probability one and if, marginally, $\mathbf{u}^n \sim \mathcal{U}([0,1)^d)$ for all $n \in 1:N$.

2.4. The Hilbert space-filling curve

The Hilbert space filling curve plays a key role in the construction and the analysis of SQMC. This curve is a Hölder continuous fractal map $H : [0,1] \to [0,1]^d$ that fills completely $[0,1]^d$; see Figure 1 for a graphical depiction, and Appendix A for a presentation of its mains properties. In what follows, we denote by $h : [0,1]^d \to [0,1]$ its pseudo-inverse which verifies, for any $\mathbf{x} \in [0,1]^d$, $H \circ h(\mathbf{x}) = \mathbf{x}$, and, for $d = 1$, we use the natural convention that H and h are the identity mappings, i.e. $H(x) = h(x) = x$, $\forall x \in [0,1]$. The Hilbert curve is not uniquely defined; in this work, we assume that H is such that $H(0) = \mathbf{0} \in [0,1]^d$ [this is in fact the classical way to construct the Hilbert curve, see e.g. 20]. This technical assumption is needed in order to be consistent with the fact that we work with left-closed and right-opened hypercubes since, in that case, $h([0,1)^d) = [0,1)$. Finally, to evaluate either h or H pointwise, one may use the algorithm of [20].

2.5. Rosenblatt transform

Another important technical tool for SQMC is the Rosenblatt transform. For a probability distribution π over $[0,1)$, F_π denotes its CDF (cumulative distribution), and F_π^{-1}

Algorithm 1 SMC Algorithm

Generate (for $n \in 1:N$) $\mathbf{x}_0^n \sim m_0(\mathrm{d}\mathbf{x}_0)$
Compute (for $n \in 1:N$) $W_0^n = G_0(\mathbf{x}_0^n)/\sum_{m=1}^N G_0(\mathbf{x}_0^m)$
for $t = 1$ to $t = T$ **do**
  Generate (for $n \in 1:N$) $u_t^n \sim \mathcal{U}[0,1)$ and set $a_t^n = F_{t-1,N}^{-1}(u_t^n)$, where $F_{t-1,N}(m) = \sum_{n=1}^N W_{t-1}^n \mathbb{1}(n \le m)$
  Generate (for $n \in 1:N$) $\mathbf{x}_t^n \sim m_t(\hat{\mathbf{x}}_{t-1}^n, \mathrm{d}\mathbf{x}_t)$, where $\hat{\mathbf{x}}_{t-1}^n = \mathbf{x}_{t-1}^{a_t^n}$
  Compute (for $n \in 1:N$) $W_t^n = G_t(\hat{\mathbf{x}}_{t-1}^n, \mathbf{x}_t^n)/\sum_{m=1}^N G_t(\hat{\mathbf{x}}_{t-1}^m, \mathbf{x}_t^m)$
end for

its inverse CDF; i.e. $F_\pi^{-1} = \inf\{x \in [0,1) : F(x) \ge u\}$. More generally, for a probability distribution π over $\mathcal{X} = [0,1)^d$, $F_\pi : \mathcal{X} \to [0,1]^d$ denotes the Rosenblatt transform, that is

$$F_\pi(\mathbf{x}) = (F_{\pi,1}(x_1), F_{\pi,2}(x_2|x_1), \dots, F_{\pi,d}(x_d|x_{1:d-1}))^T, \quad \mathbf{x} = (x_1, \dots, x_d)^T \in \mathcal{X},$$

where $F_{\pi,1}$ is the CDF of the marginal distribution of the first component (relative to π), and for $i \ge 2$, $F_{\pi,i}(\cdot|x_{1:i-1})$ is the CDF of component x_i, conditional on $(x_1, \dots, x_{i-1})$, again relative to π. (The Rosenblatt transform is not to be mistaken with the multivariate CDF; in particular it takes values in $[0,1]^d$, not in $[0,1]$.)

The inverse of F_π is denoted F_π^{-1}. The Rosenblatt transform and its inverse define a monotonous map that transforms any distribution into a uniform distribution, and vice-versa.

We overload this notation for Markov kernels: $F_{m_t}(\mathbf{x}_{t-1}, \cdot)$ is the the Rosenblatt transform of probability distribution $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ (for fixed $\mathbf{x}_{t-1} \in \mathcal{X}$), and $F_{m_t}^{-1}$ is defined similarly.

2.6. Sequential quasi-Monte Carlo

The basic structure of SMC (Sequential Monte Carlo, also known as particle filtering) algorithms is recalled as Algorithm 1. One sees from this description that SMC is a class of iterative algorithms that use resampling and mutation steps to move from a discrete approximation $\widehat{\mathbb{Q}}_t^N(\mathrm{d}\mathbf{x}_t)$ of $\mathbb{Q}_t(\mathrm{d}\mathbf{x}_t)$ to a discrete approximation $\widehat{\mathbb{Q}}_{t+1}^N(\mathrm{d}\mathbf{x}_{t+1})$ of $\mathbb{Q}_{t+1}(\mathrm{d}\mathbf{x}_{t+1})$, where

$$\widehat{\mathbb{Q}}_t^N(\mathrm{d}\mathbf{x}_t) = \sum_{n=1}^N W_t^n \delta_{\mathbf{x}_t^n}(\mathrm{d}\mathbf{x}_t), \quad t \in 0:T.$$

A closer look at Algorithm 1 shows that, for $t \ge 1$, the resampling and the mutation steps together amounts to sampling from the (random) distribution on $\mathcal{X}^2$ defined by

$$\pi_t^N(\mathrm{d}(\mathbf{x}_{t-1}, \mathbf{x}_t)) = \widehat{\mathbb{Q}}_{t-1}^N \otimes m_t(\mathrm{d}(\mathbf{x}_{t-1}, \mathbf{x}_t)) \tag{1}$$

where, for a probability measure $\pi \in \mathcal{P}([0,1)^{d_1})$ and a kernel $K : [0,1)^{d_1} \to \mathcal{P}([0,1)^{d_2})$, the notation $\pi \otimes K(\mathrm{d}(\mathbf{x}_1, \mathbf{x}_2))$ denotes the probability measure $\pi(\mathrm{d}\mathbf{x}_1)K(\mathbf{x}_1, \mathrm{d}\mathbf{x}_2)$ on $[0,1)^{d_1+d_2}$.

Based on this observation, the basic idea of SQMC is to replace the uniform random numbers used at iteration $t \geq 1$ of an SMC algorithm to sample from (1) by a QMC point set $\mathbf{u}_t^{1:N}$ of appropriate dimension. In the deterministic version of SQMC, the only known property of $\mathbf{u}_t^{1:N}$ is that its discrepancy converges to zero as N goes to infinity. Thus, we must make sure that the transform applied to $\mathbf{u}_t^{1:N}$ preserves consistency (relative the extreme norm): i.e. $D(\mathbf{u}^{1:N}) \to 0$ implies that $\|\Gamma_t^N(\mathbf{u}^{1:N}) - \pi_t^N\|_{\mathrm{E}} \to 0$, where Γ_t^N is the chosen transformation.

When the state space is univariate, [17] propose to use for Γ_t^N the inverse Rosenblatt transformation of π_t^N described in the previous subsection, which amounts to sample $(\hat{x}_{t-1}^n, x_t^n)$ from (1) as follows:

$$\hat{x}_{t-1}^n = F^{-1}_{\widehat{\mathbb{Q}}_{t-1}^N}(u_t^n), \quad x_t^n = F^{-1}_{m_t}(\hat{x}_{t-1}, v_t^n), \quad (u_t^n, v_t^n) \sim \mathcal{U}([0,1)^2).$$

However, when the state variable is multivariate (i.e. $d > 1$) this approach cannot be directly used because in that case $\widehat{\mathbb{Q}}_{t-1}^N(\mathrm{d}\mathbf{x}_{t-1})$ is a weighted sum of Dirac measures over $\mathcal{X} \subset \mathbb{R}^d$.

To extend this approach to multidimensional state-space models, [17] transform the multivariate distribution $\widehat{\mathbb{Q}}_{t-1}^N(\mathrm{d}\mathbf{x}_{t-1})$ into a univariate distribution $\widehat{\mathbb{Q}}_{t-1,h}^N(\mathrm{d}h_{t-1})$ by applying the change of variable $h : \mathcal{X} \to [0,1)$, where h is the pseudo-inverse of the Hilbert curve (see Section 3.4). Using this change of variable, the resampling and mutation steps of SMC are equivalent to sampling from

$$\pi_{t,h}^N(\mathrm{d}(h_{t-1}, \mathbf{x}_t)) = \widehat{\mathbb{Q}}_{t-1,h}^N \otimes m_{t,h}(\mathrm{d}(h_{t-1}, \mathbf{x}_t)) \tag{2}$$

where $m_{t,h}(h_{t-1}, \mathbf{x}_t) := m_t(H(h_{t-1}), \mathbf{x}_t)$. As for the univariate setting, one can generate random variates from $\pi_{t,h}^N(\mathrm{d}(h_{t-1}, \mathbf{x}_t))$ using the inverse Rosenblatt transformation of this distribution; that is, we can sample $(\hat{h}_{t-1}^n, \mathbf{x}_t^n)$ from (2) as follows:

$$\hat{h}_{t-1}^n = F^{-1}_{\widehat{\mathbb{Q}}_{t-1,h}^N}(u_t), \quad \mathbf{x}_t^n = F^{-1}_{m_t}(H(\hat{h}_{t-1}^n), \mathbf{v}_t^n), \quad (u_t^n, \mathbf{v}_t^n) \sim \mathcal{U}([0,1)^{d+1})$$

where $\widehat{\mathbb{Q}}_{t-1,h}^N = \sum_{n=1}^N W_{t-1}^n \delta_{h(\mathbf{x}_{t-1}^n)}(\mathrm{d}h_t)$. Since the $h(\mathbf{x}_{t-1}^n)$'s lie in $[0,1]$, computing $\hat{h}_{t-1}$ above amounts to (a) sort the $h(\mathbf{x}_{t-1}^n)$, i.e. find permutation σ_{t-1} such that $h(\mathbf{x}_{t-1}^{\sigma_{t-1}(1)}) \leq \ldots \leq h(\mathbf{x}_{t-1}^{\sigma_{t-1}(N)})$; and (b) for each n find the integer a_{t-1}^n such that $\sum_{i<a_{t-1}^n} W_{t-1}^{\sigma_{t-1}(i)} < u_t^n \leq \sum_{i \leq a_{t-1}} W_{t-1}^{\sigma_{t-1}(i)}$. (Provided the u_t^n are ordered, Step (b) may be performed in $\mathcal{O}(N)$ time.)

The resulting SQMC algorithm, which is therefore based for $t \geq 1$ on $d+1$-dimensional QMC point sets $\mathbf{u}_t^{1:N}$, $\mathbf{u}_t^n = (u_t^n, \mathbf{v}_t^n) \in [0,1)^{d+1}$, is presented in Algorithm 2.

Note that, before [17], the idea of introducing $T+1$ point sets (one per time step) of dimension d, in order to perform integration with respect to a space of dimension $(T+1)d$ may be found in the array-RQMC algorithm [24, 15, 25, 23], which is designed

Algorithm 2 SQMC Algorithm

Generate a QMC point set $\mathbf{u}_0^{1:N}$ in $[0,1)^d$
Compute (for $n \in 1:N$) $\mathbf{x}_0^n = F_{m_0}^{-1}(\mathbf{u}_0^n)$
Compute (for $n \in 1:N$) $W_0^n = G_0(\mathbf{x}_0^n)/\sum_{m=1}^N G_0(\mathbf{x}_0^m)$
for $t = 1$ to $t = T$ **do**
Generate a QMC point set $\mathbf{u}_t^{1:N}$ in $[0,1)^{d+1}$, let $\mathbf{u}_t^n = (u_t^n, \mathbf{v}_t^n)$, where $u_t^n \in [0,1)$, $\mathbf{v}_t^n \in [0,1)^d$. Assume that, for all $n, m \in 1:N$, $n \le m \implies u_t^n \le u_t^m$
Hilbert sort: find permutation σ_{t-1} such that $h(\mathbf{x}_{t-1}^{\sigma_{t-1}(1)}) \le \ldots \le h(\mathbf{x}_{t-1}^{\sigma_{t-1}(N)})$
Compute (for $n \in 1:N$) $a_{t-1}^n = F_{t-1,N}^{-1}(u_t^n)$ where $F_{t-1,N}(m) = \sum_{i=1}^N W_{t-1}^{\sigma_{t-1}(i)} \mathbb{I}(i \le m)$
Compute (for $n \in 1:N$) $\mathbf{x}_t^n = F_{m_t}^{-1}(\hat{\mathbf{x}}_{t-1}^n, \mathbf{v}_t^n)$, where $\hat{\mathbf{x}}_{t-1}^n = \mathbf{x}_{t-1}^{a_{t-1}^n}$
Compute (for $n \in 1:N$) $W_t^n = G_t(\hat{\mathbf{x}}_{t-1}^n, \mathbf{x}_t^n)/\sum_{m=1}^N G_t(\hat{\mathbf{x}}_{t-1}^m, \mathbf{x}_t^m)$
end for

to evaluate expectations with respect to a Markov chain $(\mathbf{x}_t)$, run from $t = 0$ to time T. The array-RQMC algorithm requires to specify a certain order on the state space $\mathcal{X}$, and the Hilbert curve used in SQMC may be seen as one particular way to order points in $\mathcal{X}$ [37]. On the other hand, the theoretical results of [17] and of this paper rely quite heavily on properties of the Hilbert curve, which seems to indicate that the Hilbert curve is the 'right' way to sort ancestors in SQMC.

3. Preliminary results

3.1. Consistency of SQMC

The consistency of Algorithm 2 (as $N \to +\infty$, with respect to the extreme metric) was established in [17, Theorem 5], under the assumption that F_{m_t} is Lipschitz. We generalise below this result to the case where F_{m_t} is Hölder continuous, as this generalisation will be needed later on when dealing with the backward step. This also allows us to recall some of the assumptions that will be repeated throughout the paper. For convenience, let $F_{m_t}(\mathbf{x}_{t-1}, \mathbf{x}_t) = F_{m_0}(\mathbf{x}_0)$ when $t = 0$.

Theorem 1. *Consider the set-up of Algorithm 2 where, for all $t \in 0:T$, $(\mathbf{u}_t^{1:N})_{N\ge1}$ is a sequence of point sets in $[0,1)^{d_t}$, with $d_0 = d$ and $d_t = d+1$ for $t > 0$, such that $D(\mathbf{u}_t^{1:N}) \to 0$ as $N \to +\infty$. Assume the following holds for all $t \in 0:T$:*

1. *The $\mathbf{x}_t^n$'s are pairwise distinct: $\mathbf{x}_t^n \ne \mathbf{x}_t^m$ for $n \ne m \in 1:N$;*
2. *G_t is continuous and bounded;*
3. *$F_{m_t}(\mathbf{x}_{t-1}, \mathbf{x}_t)$ is such that*
   a) *For $i \in 1:d$ and for a fixed $\mathbf{x}'$, the i-th coordinate of $F_{m_t}(\mathbf{x}', \mathbf{x})$ is strictly increasing in $x_i \in [0,1)$, the i-th coordinate of $\mathbf{x}$;*

b) Viewed as a function of $\mathbf{x}'$ and $\mathbf{x}$, $F_{m_t}(\mathbf{x}', \mathbf{x})$ is Hölder continuous;

c) For $i \in 1:d$, $m_{ti}(\mathbf{x}', x_{1:i-1}, \mathrm{d}x_i)$, the distribution of the component x_i conditional on $(x_1, .., x_{i-1})$ relative to $m_t(\mathbf{x}', \mathrm{d}\mathbf{x})$, admits a density $p_{ti}(x_i|x_{1:i-1}, \mathbf{x}')$ with respect to the Lebesgue measure such that $\|p_{ti}(\cdot|\cdot)\|_\infty < +\infty$.

4. $\mathbb{Q}_t(\mathrm{d}\mathbf{x}_t) = p_t(\mathbf{x}_t)\lambda_d(\mathrm{d}\mathbf{x}_t)$ where $p_t(\mathbf{x}_t)$ is a strictly positive bounded density.

For $t \in 1:T$, let $P^N_{t,h} = (h(\hat{\mathbf{x}}^{1:N}_{t-1}), \mathbf{x}^{1:N}_t)$. Then, under Assumptions 1-4, we have, for $t \in 1:T$,

$$\left\|\mathcal{S}(P^N_{t,h}) - \mathbb{Q}_{t-1,h} \otimes m_{t,h}\right\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty$$

and, for $t \in 0:T$,

$$\|\widehat{\mathbb{Q}}^N_t - \mathbb{Q}_t\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty.$$

Theorem 1 applies to either deterministic or random point sets $\mathbf{u}^{1:N}_t$. In the latter case, assumptions relative to the $\mathbf{u}^{1:N}_t$'s must hold almost surely.

The difference with [17, Theorem 5] is Assumption 3, where 3c was not needed but it was assumed that F_{m_t} is a Lipschitz function. In this work, Assumption 3c will be required to establish the validity of the backward step. Assumption 1 is a technical condition that is verified almost surely for the randomized version of SQMC while assuming that G_t is bounded is standard in particle filtering [9].

The proof of Theorem 1 is omitted since it can be directly deduced from the proof of [17, Theorem 5] and from the generalization of the result of [22, "Satz 2"] presented in the Section 3.3, which constitutes one of the key ingredients to study the backward pass of SQMC.

3.2. Backward decomposition

Backward smoothing algorithms require that the Markov kernel $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ admits a (strictly positive) probability density which may be computed pointwise; $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t) = m_t(\mathbf{x}_{t-1}, \mathbf{x}_t)\lambda_d(\mathrm{d}\mathbf{x}_t)$, with $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t) > 0$ (and λ_d being the Lebesgue measure in our case).

The backward decomposition of the smoothing distribution is [e.g. 10]:

$$\widetilde{\mathbb{Q}}_T(\mathrm{d}\mathbf{x}_{0:T}) = \mathbb{Q}_T(\mathrm{d}\mathbf{x}_T) \prod_{t=1}^{T} \mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) \tag{3}$$

where, for any $\pi \in \mathcal{P}(\mathcal{X})$ and $t \in 1:T$, $\mathcal{M}_{t,\pi} : \mathcal{X} \to \mathcal{P}(\mathcal{X})$ is the Markov kernel such that

$$\mathcal{M}_{t,\pi}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) := \tilde{G}_t(\mathbf{x}_{t-1}, \mathbf{x}_t)\pi(\mathrm{d}\mathbf{x}_{t-1})$$

with

$$\tilde{G}_t(\mathbf{x}_{t-1}, \mathbf{x}_t) := \frac{m_t(\mathbf{x}_{t-1}, \mathbf{x}_t)G_t(\mathbf{x}_{t-1}, \mathbf{x}_t)}{\int_{\mathcal{X}} m_t(\tilde{\mathbf{x}}_{t-1}, \mathbf{x}_t)G_t(\tilde{\mathbf{x}}_{t-1}, \mathbf{x}_t)\pi(\mathrm{d}\tilde{\mathbf{x}}_{t-1})}. \tag{4}$$

As a preliminary result, we show that the plug-in estimate $\widetilde{\mathbb{Q}}^N_T$ of $\widetilde{\mathbb{Q}}_T$, obtained by replacing $\mathbb{Q}_t$ with $\widehat{\mathbb{Q}}^N_t$ in (3), is consistent for the extreme norm; see Appendix B.1 for a proof.

Theorem 2. *Consider the set-up of Algorithm 2, define for* $t \in 1:T$

$$\widetilde{\mathbb{Q}}_t^N(\mathrm{d}\mathbf{x}_{0:t}) = \widehat{\mathbb{Q}}_t^N(\mathrm{d}\mathbf{x}_t)\prod_{s=1}^{t}\mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1}), \tag{5}$$

and consider the following hypotheses:

(H1) $\tilde{G}_t$ *is continuous and bounded,* $\|\tilde{G}_t\|_\infty < \infty$*;*

(H2) $F_{\mathcal{M}_{t,\mathbb{Q}_{t-1}}}(\mathbf{x}_t, \mathbf{x}_{t-1})$ *satisfies Assumptions 3a and 3b of Theorem 1 (i.e. replace* m_t *by* $\mathcal{M}_{t,\mathbb{Q}_{t-1}}$ *in these assumptions).*

Then,

1. *Under (H1) and the assumptions of Theorem 1, one has (for* $t \in 1:T$*)*

$$\sup_{\mathbf{x}_t\in[0,1)^d} \|\mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) - \mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1})\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty. \tag{6}$$

2. *If (6) holds, and under (H2) and the assumptions of Theorem 1, one has (for* $t \in 1:T$*)*

$$\|\widetilde{\mathbb{Q}}_t^N - \widetilde{\mathbb{Q}}_t\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty. \tag{7}$$

The first result above does not have a clear interpretation, but it will be used as an intermediate result later on.

3.3. A generalization of Satz 2 of [22]

Theorem 3 below generalizes Proposition 'Satz 2' of [22] to the case where point sets in $[0,1)^d$ are transformed through a Hölder continuous Rosenblatt transformation; see Appendix B.2 for a proof.

Theorem 3. *Let* π *be a probability measure on* $[0,1)^d$ *and assume the following:*

1. *Viewed as a function of* $\mathbf{x}$*,* $F_\pi(\mathbf{x})$ *is Hölder continuous with Hölder exponent* $\kappa \in (0,1]$*;*

2. *For* $i \in 1:d$*, the* i*-th coordinate of* $F_\pi(\mathbf{x})$ *is strictly increasing in* $x_i \in [0,1)$*, the* i*-th coordinate of* $\mathbf{x}$*;*

3. *For* $i \in 1:d$*,* $\pi_i(x_{1:i-1}, \mathrm{d}x_i)$*, the distribution of the component* x_i *conditional on* $(x_1,..,x_{i-1})$ *relative to* $\pi(\mathrm{d}\mathbf{x})$*, admits a density* $p_i(x_i|x_{1:i-1})$ *with respect to the Lebesgue measure such that* $\|p_i(\cdot|\cdot)\|_\infty < +\infty$*.*

Let $\mathbf{u}^{1:N}$ *be a point set in* $[0,1)^d$ *and, for* $n \in 1:N$*, define* $\mathbf{x}^n = F_\pi^{-1}(\mathbf{u}^n)$*. Then, for a constant* $c > 0$*,*

$$\|\mathcal{S}(\mathbf{x}^{1:N}) - \pi\|_{\mathrm{E}} \le cD(\mathbf{u}^{1:N})^{1/\tilde{d}}$$

where $\tilde{d} = \sum_{i=0}^{d-1}\lceil\kappa^{-1}\rceil^i$*.*

When the Rosenblatt transformation F_π is Lipschitz, $\tilde{d}=d$ and we recover the result of [22]. In this case, Assumption 3 is not needed. Notice that the rate provided in Theorem 3 decreases quickly with the Hölder exponent κ. For $\kappa=1/2$, the convergence rate is of order $\mathcal{O}(D(\mathbf{u}^{1:N})^{1/(2^d-1)})$ and hence is very slow even for moderate values of d.

We will see that the backward step of the forward-backward SQMC smoothing algorithm amounts to applying to QMC point sets transformations that are "nearly" $(1/d)$-Hölder continuous (in a sense that we will make precise). The main message of Theorem 3, as far as SQMC is concerned, is that such an algorithm may be consistent (as $N\rightarrow+\infty$) despite being based on non-Lipschitz transformations.

Theorem 3 is interesting more generally, since the construction of low discrepancy point sets with respect to non uniform distributions is an important problem, which is motivated by the generalized Koksma-Hlawka inequality [2, Theorem 1]:

$$\left|\frac{1}{N}\sum_{n=1}^{N}\varphi(\mathbf{x}^n)-\int_{[0,1)}\varphi(\mathbf{x})\pi(\mathrm{d}\mathbf{x})\right|\leq V(\varphi)\|\mathcal{S}(\mathbf{x}^{1:N})-\pi\|_{\mathrm{E}}$$

where $V(\varphi)$ is the variation of φ in the sense of Hardy and Krause. It is also interesting to mention that the inverse Rosenblatt transformation is the best known construction of low discrepancy point sets for non uniform probability measures, although the bounds for the extreme metric given in [22, "Satz 2"] and in Theorem 3 are very far from the best known achievable rate since [1, Theorem 1] have established the existence, for any probability measure π on $[0,1)^d$, of a sequence $(\mathbf{x}^n)_{n\geq1}$ verifying $\|\mathcal{S}(\mathbf{x}^{1:N})-\pi\|_{\mathrm{E}}=\mathcal{O}(N^{-1}(\log N)^{0.5(3d+1)})$.

3.4. Discrepancy conversion through the Hilbert space filling curve

We now state results regarding how the Hilbert curve $H:[0,1]\rightarrow[0,1]^d$ conserves discrepancy. Such results were not directly needed to establish the consistency of SQMC. Indeed, as outlined in the statement of Theorem 1, it was sufficient to show that $P_{t,h}^N$ has low discrepancy with respect to the proposal distribution $\mathbb{Q}_{t-1,h}\otimes m_{t,h}$, where we recall that $P_{t,h}^N=\big(h(\hat{\mathbf{x}}_{t-1}^{1:N}),\mathbf{x}_t^{1:N}\big)$, with $h(\hat{\mathbf{x}}_{t-1}^{1:N})\in[0,1)$. The discrepancy of the "resampled" particles $\hat{\mathbf{x}}_{t-1}^{1:N}$ in $[0,1)^d$ was not derived. But, again, we will need such results when dealing with backward estimates.

More precisely, and as explained below (see Section 5.2), the analysis of backward estimates require results on the conversion of discrepancies through the following mapping, defined for $k\in\mathbb{N}$, by

$$H_k:(x_0,\dots,x_k)\in[0,1)^{(k+1)}\mapsto(H(x_0),\dots,H(x_k))\in[0,1)^{d(k+1)}\tag{8}$$

and with pseudo-inverse $h_k:[0,1)^{d(k+1)}\rightarrow[0,1)^{k+1}$.

Theorem 4 and Corollary 1 below are generalizations of [35, Theorem 1], which corresponds to Theorem 4 with $k=0$, π_h the uniform distribution on $[0,1)$ and $\pi_h^N=\mathcal{S}(u^{1:N})$ for a point set $u^{1:N}$ in $[0,1)$. To save space, the proofs the these two results are omitted.

Theorem 4. *Let* $\pi(\mathrm{d}\mathbf{x}) = \pi(\mathbf{x})\lambda_{d(k+1)}(\mathrm{d}\mathbf{x})$*,* $k \in \mathbb{N}$*, be a probability measure on* $[0,1)^{d(k+1)}$ *with bounded density* π*,* π_{h_k} *be the image of* π *by* h_k *and* $(\pi^N_{h_k})_{N\geq 1}$ *be a sequence of probability measures on* $[0,1)^{k+1}$ *such that* $\|\pi^N_{h_k} - \pi_{h_k}\|_{\mathrm{E}} \to 0$ *as* $N \to +\infty$*. Let* π^N *be the image by* H_k *of* $\pi^N_{h_k}$*. Then,*

$$\|\pi^N - \pi\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty.$$

Corollary 1. *Consider the set-up of Theorem 4 with* $k = 0$ *and let* $K : [0,1)^d \to \mathcal{P}([0,1)^s)$ *be a Markov kernel,* $K_h(h_1, \mathrm{d}\mathbf{x}_2) = K(h(\mathbf{x}_1), \mathrm{d}\mathbf{x}_2)$ *and* $P^N_h = (h_1^{1:N}, \mathbf{x}_2^{1:N})$ *be a sequence of point sets in* $[0,1)^{1+s}$ *such that, as* $N \to +\infty$*,* $\|\mathcal{S}(P^N_h) - \pi_h \otimes K_h\|_{\mathrm{E}} \to 0$*. Let* $P^N = \left(H(h_1^{1:N}), \mathbf{x}_2^{1:N}\right)$*. Then,*

$$\|\mathcal{S}(P^N) - \pi^N \otimes K\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty.$$

A direct consequence of this corollary is that, under the assumptions of Theorem 1, the point set $P^N_t = (\hat{\mathbf{x}}^{1:N}_{t-1}, \mathbf{x}^{1:N}_t)$ is such that, as $N \to +\infty$,

$$\|\mathcal{S}(P^N_t) - \mathbb{Q}_{t-1} \otimes m_t\|_{\mathrm{E}} \to 0.$$

Another consequence of this corollary is that Algorithm 2 can be trivially adapted to forward smoothing, as briefly explained in the next section.

4. SQMC forward smoothing

Consider now the following extension of Algorithm 2, where full trajectories $\mathbf{z}_t := \mathbf{x}_{0:t} \in \mathcal{X}^{t+1}$ are carried forward: at time 0, set $\mathbf{z}^n_0 := \mathbf{x}^n_0$, and, recursively, $\mathbf{z}^n_t := (\hat{\mathbf{z}}^n_t, \mathbf{x}^n_t)$, with $\hat{\mathbf{z}}^n_t := \mathbf{z}^{a^n_{t-1}}_{t-1}$. In addition, replace the Hilbert sort step of Algorithm 2 by the same operation on full trajectories:

Hilbert sort: find permutation σ_{t-1} such that $h^t(\mathbf{z}^{\sigma_{t-1}(1)}_{t-1}) \leq \ldots \leq h^t(\mathbf{z}^{\sigma_{t-1}(N)}_{t-1})$

with h^t the inverse of a Hilbert curve H^t that maps $[0,1]$ into $[0,1]^{dt}$. In other words, this is the SQMC equivalent of the smoothing technique known as 'forward smoothing'.

Proposition 1. *Under Assumptions 1-3 of Theorem 1, and Assumption 4'*

4'. $\widetilde{\mathbb{Q}}_t(\mathrm{d}\mathbf{z}_t) = \tilde{p}_t(\mathbf{z}_t)\lambda_{d(t+1)}(\mathrm{d}\mathbf{z}_t)$ where $\tilde{p}_t(\mathbf{z}_t)$ is a strictly positive bounded density;

one has, for $t \geq 0$ *and the forward smoothing algorithm described above,*

$$\left\|\sum_{n=1}^N W^n_t \delta_{\mathbf{z}^n_t} - \widetilde{\mathbb{Q}}_t\right\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty \tag{9}$$

where $\widetilde{\mathbb{Q}}_t$ *denotes the smoothing distribution at time* t*.*

See Appendix B.3 for a proof.

This result is presented for the sake of completeness, but it is clear that it is of limited practical interest. Transformations through H^t will lead to poor convergence rates as soon as t becomes large, as per Theorem 4. In addition, there is no reason to believe that the SQMC version of forward smoothing would not suffer from the same major drawback as its Monte Carlo counterpart, namely that the N simulated paths quickly coalesce to a single ancestor.

5. SQMC backward smoothing

We now turn to the derivation and analysis of a SQMC version of backward smoothing. There exist in fact two backward smoothing algorithms. The first one [14] approximates the marginal smoothing distributions $\mathbb{Q}_{t|T}(\mathrm{d}\mathbf{x}_t)$ for $t \in 0:T$; that is, the marginal distribution of $\mathbf{x}_t$ relative to $\widetilde{\mathbb{Q}}_T(\mathrm{d}\mathbf{x}_{0:T})$. This may be used to compute the smoothing expectation of additive functions $\varphi(\mathbf{x}_{0:T}) = \sum_{t=0}^T \varphi_t(\mathbf{x}_t)$ such as, e.g., the score functions of certain models [e.g. 33]. See Section 5.1.

The second type of backward step [18] allows to estimate the full (joint) smoothing distribution $\widetilde{\mathbb{Q}}(\mathrm{d}\mathbf{x}_{0:T})$. Its SQMC version is given and analysed in Section 5.2.

These two algorithms share the following properties: (a) they require that the Markov kernel $m_t(\mathbf{x}_{t-1}, \mathrm{d}\mathbf{x}_t)$ admits a positive probability density $m_t(\mathbf{x}_{t-1}, \mathbf{x}_t)$ which may be computed pointwise (for all $\mathbf{x}_{t-1}, \mathbf{x}_t \in \mathcal{X}$); (b) they use as input the output of a forward pass, i.e. either Algorithm 1 (SMC), or Algorithm 2 (SQMC); and (c) their complexity is $\mathcal{O}(TN^2)$.

5.1. Marginal backward smoothing

To perform marginal smoothing, one computes, from the output of the forward pass, the following smoothing weights:

$$\widetilde{W}_{t|T}^n := W_t^n \times \sum_{m=1}^N \frac{\widetilde{W}_{t+1|T}^m m_{t+1}(\mathbf{x}_t^n, \mathbf{x}_{t+1}^m) G_{t+1}(\mathbf{x}_t^n, \mathbf{x}_{t+1}^m)}{\sum_{p=1}^N W_t^p m_{t+1}(\mathbf{x}_t^p, \mathbf{x}_{t+1}^m) G_{t+1}(\mathbf{x}_t^p, \mathbf{x}_{t+1}^m)}$$

for all $n \in 1:N$, and recursively, from $t = T-1$, to $t = 0$. For $t = T$, simply set $\widetilde{W}_{t|T}^n = W_T^n$. These weights are obtained by marginalising recursively the joint distribution given in (5) over $\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_{t+1}$. One has:

$$\widetilde{\mathbb{Q}}_{t|T}^N(\mathrm{d}\mathbf{x}_t) := \sum_{n=1}^N \widetilde{W}_{t|T}^n \delta_{\mathbf{x}_t^n}(\mathrm{d}\mathbf{x}_t) \approx \mathbb{Q}_{t|T}(\mathrm{d}\mathbf{x}_t).$$

This particular backward pass may be applied to either the output of SMC (Algorithm 1), or SQMC (Algorithm 2). In the latter case, the question is whether this approach remains valid. The answer is directly given by Theorem 2: under its assumptions, we have that

$$\|\widetilde{\mathbb{Q}}_{t|T}^N - \widetilde{\mathbb{Q}}_{t|T}\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty$$

since $\widetilde{\mathbb{Q}}_{t|T}$ (resp. $\widetilde{\mathbb{Q}}^N_{t|T}$) is a certain marginal distribution of $\widetilde{\mathbb{Q}}_T$ (resp. $\widetilde{\mathbb{Q}}^N_T$). In words, marginal backward smoothing generates consistent (marginal) smoothing estimates when applied to the output of the SQMC algorithm.

5.2. Full backward smoothing

The SQMC backward step to estimate the joint smoothing distribution $\widetilde{\mathbb{Q}}_T$, proposed in [17], is recalled as Algorithm 3.

Algorithm 3 SQMC Backward step for full smoothing

Input: $\mathbf{x}_t^{\sigma_t(1:N)}$, $W_t^{\sigma_t(1:N)}$ for $t \in 0:T$, output of Algorithm 2 after the Hilbert sort step (i.e, for all $n, m \in 1:N$, $n \leq m \implies h(\mathbf{x}_t^{\sigma_t(n)}) \leq h(\mathbf{x}_t^{\sigma_t(m)})$) and $\tilde{\mathbf{u}}^{1:N}$ a point set in $[0,1)^{T+1}$ such that, for all $n, m \in 1:N$, $n \leq m \implies u_T^n \leq u_T^m$

Output: $\tilde{\mathbf{x}}_{0:T}^{1:N}$ (N trajectories in $\mathcal{X}^{T+1}$)

for $n = 1 \to N$ **do**

Compute $\tilde{\mathbf{x}}_T^n = \mathbf{x}_T^{a_T^n}$ where $a_T^n = F_{T,N}^{-1}(u_T^n)$ with $F_{T,N}(i) = \sum_{m=1}^N W_T^{\sigma_T(m)} \mathbb{I}(m \leq i)$

end for

for $t = T-1 \to 0$ **do**

for $n = 1 \to N$ **do**

Compute $\tilde{\mathbf{x}}_t^n = \mathbf{x}_t^{\tilde{a}_t^n}$ where $\tilde{a}_t^n = \tilde{F}_{t,N}^{-1}(\tilde{\mathbf{x}}_{1+1}^n, \tilde{u}_t^n)$ with $\tilde{F}_{t,N}(\mathbf{x}_{t+1}, i) = \sum_{m=1}^N \widetilde{W}_t^{\sigma_t(m)}(\mathbf{x}_{t+1}) \mathbb{I}(m \leq i)$, and $\widetilde{W}_t^m(\mathbf{x}_{t+1}) = \frac{W_t^m m_{t+1}(\mathbf{x}_t^m, \mathbf{x}_{t+1}) G_{t+1}(\mathbf{x}_t^m, \mathbf{x}_{t+1})}{\sum_{p=1}^N W_t^p m_{t+1}(\mathbf{x}_t^p, \mathbf{x}_{t+1}) G_{t+1}(\mathbf{x}_t^p, \mathbf{x}_{t+1})}$.

end for

end for

Algorithm 3 generates a low discrepancy point set for distribution $\widetilde{\mathbb{Q}}_T^N$, the plug-in estimate of $\widetilde{\mathbb{Q}}_T$, and is therefore the exact QMC equivalent of the backward step of standard backward sampling.

To better understand why Algorithm 3 is valid, it helps to decompose it in two steps. First, it transforms $\tilde{\mathbf{u}}^{1:N}$, a point set in $[0,1)^{T+1}$, into $\tilde{h}_{0:T}^{1:N}$, another point set in $[0,1)^{T+1}$, by applying the inverse Rosenblatt transformation of

$$\widetilde{\mathbb{Q}}_{T,h}^N(\mathrm{d}h_{0:T}) := \widehat{\mathbb{Q}}_{T,h}^N(\mathrm{d}h_T) \prod_{t=1}^T \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1,h}^N}^h(h_t, \mathrm{d}h_{t-1}), \tag{10}$$

which is the image of probability measure $\widetilde{\mathbb{Q}}_T^N(\mathrm{d}\mathbf{x}_{0:T})$, defined in (5), by mapping $h_T : (\mathbf{x}_0, \ldots, \mathbf{x}_T) \mapsto (h(\mathbf{x}_0), \ldots, h(\mathbf{x}_T))$. Recall that $\widehat{\mathbb{Q}}_{t,h}^N$ is the image of $\widehat{\mathbb{Q}}_t^N$ by h while, for any $\pi \in \mathcal{P}([0,1))$ and $t \in 1:T$, $\mathcal{M}_{t+1,\pi}^h : [0,1) \to \mathcal{P}([0,1))$ is a Markov kernel such that

$$\mathcal{M}_{t,\pi}^h(h_t, \mathrm{d}h_{t-1}) \propto m_t\big(H(h_{t-1}), H(h_t)\big) G_t\big(H(h_{t-1}), H(h_t)\big) \pi(\mathrm{d}h_{t-1}).$$

In a second step, Algorithm 3 returns $\tilde{\mathbf{x}}_{0:T}^{1:N}$ where $\tilde{\mathbf{x}}_{0:T}^n = H_T(\tilde{h}_{0:T}^n)$ with the mapping $H_T : [0,1)^{T+1} \to [0,1)^{d(T+1)}$ defined in (8).

5.2.1. L_1- and L_2-convergence

A direct consequence of the inverse Rosenblatt interpretation of the previous section is that, when Algorithm 3 uses a RQMC point set as input, the random point $\tilde{\mathbf{x}}^n_{0:T}$ is such that, for any function $\varphi : [0,1)^{d(T+1)} \to \mathbb{R}$ and for any $n \in 1:N$, we have $\mathbb{E}[\varphi(\tilde{\mathbf{x}}^n_{0:T})|\mathcal{F}^N_T] = \widetilde{\mathbb{Q}}^N_T(\varphi)$, with $\mathcal{F}^N_T$ the σ-algebra generated by the forward step. Together with Theorem 2, this observation allows us to deduce L_2-convergence for test functions φ that are continuous and bounded (see Appendix B.4 for a proof).

Theorem 5. *Consider the set-up of the SQMC forward filtering-backward smoothing algorithm (Algorithms 2 and 3) and assume the following:*

1. *In Algorithm 2, $(\mathbf{u}^{1:N}_t)_{N\geq 1}$, $t \in 0:T$, are independent random sequences of point sets in $[0,1)^{d_t}$, with $d_0 = d$ and $d_t = d+1$ for $t>0$, such that, for any $\epsilon > 0$, there exists a $N_{\epsilon,t} > 0$ such that, almost surely, $D(\mathbf{u}^{1:N}_t) \leq \epsilon$, $\forall N \geq N_{\epsilon,t}$;*
2. *In Algorithm 3, $(\tilde{\mathbf{u}}^{1:N})_{N\geq 1}$ is a sequence of point sets in $[0,1)^{T+1}$ such that*
   a) *$\forall n \in 1:N$, $\tilde{\mathbf{u}}^n \sim \mathcal{U}([0,1)^{T+1})$;*
   b) *For any function $\varphi \in L_2\left([0,1)^{d(T+1)}, \lambda_{d_t}\right)$, $\mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^N \varphi(\mathbf{u}^n_t)\right) \leq C\sigma^2_\varphi r(N)$ where $\sigma^2_\varphi = \int \left\{\varphi(\mathbf{u}) - \int \varphi(\mathbf{v})\mathrm{d}\mathbf{v}\right\}^2 \mathrm{d}\mathbf{u}$, $r(N) \to 0$ as $N \to +\infty$, and where both C and $r(N)$ do not depend on φ;*
3. *Assumptions of Theorem 1 and Assumptions H1-H2 of Theorem 2 hold.*

Then, for any continuous and bounded function $\varphi : \mathcal{X}^{T+1} \to \mathbb{R}$,

$$\mathbb{E}\left|\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi) - \widetilde{\mathbb{Q}}_T(\varphi)\right| \to 0, \quad \mathrm{Var}\left(\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi)\right) \to 0, \quad \text{as } N \to +\infty.$$

Assumption 1 is verified for instance when $\mathbf{u}^{1:N}_t$ consists of the first N points of a nested scrambled (t, d_t)-sequence in base $b \geq 2$ [30, 31, 32]. The result above may be easily extended to the case where the $\mathbf{u}^{1:N}_t$'s are deterministic (rather than random) QMC point sets.

On the other hand, the point set $\tilde{\mathbf{u}}^{1:N}$ used as input of the backward pass is necessarily random (for the result above to hold). But $\tilde{\mathbf{u}}^{1:N}$ does not need to be a QMC point set (i.e. to have low discrepancy). In particular, Assumption 2 is satisfied when the $\tilde{\mathbf{u}}^{1:N}$ are IID uniform variates (in $[0,1)^{T+1}$); then $C=1$ and $r(N) = N^{-1}$. See Section 5.3 for a discussion on the use of QMC or pseudo-random numbers in the backward step of SQMC.

5.2.2. Consistency

Compared to standard (forward) SQMC, establishing the consistency of SQMC backward smoothing requires two extra technical steps. First, as Algorithm 3 generates a point set $\tilde{h}^{1:N}_{0:T}$ in $[0,1)^{T+1}$ using the inverse Rosenblatt transformation of the probability measure defined in (10), and then projects it back to $\mathcal{X}^{T+1}$ through H_T, we need to establish that this transformation preserves the low discrepancy properties of $\tilde{h}_{0:T}^{1:N}$. For this we will use Theorem 4.

Second, the proof of [17] for the consistency of SQMC required smoothness conditions on the Rosenblatt transformation of $m_{t,h}(h_{t-1}, \mathrm{d}\mathbf{x}_t) = m_t(H(h_{t-1}), \mathrm{d}\mathbf{x}_t)$, so that this transformation maintains low discrepancy, as explained in Section 2.6. Due to the Hölder property of the Hilbert curve, the Hölder continuity of F_{m_t} implies that $F_{m_{t,h}}$ is Hölder continuous as well. Similarly, for the backward step we need assumptions on the Markov kernel $\mathcal{M}_{t,\mathbb{Q}_{t-1}}$ which imply sufficient smoothness for the Rosenblatt transformation of $\mathcal{M}^h_{t,\widehat{\mathbb{Q}}^N_{t-1,h}}$ which is used in the course of Algorithm 3 to transform the QMC point set in $[0,1)^{T+1}$.

To this aim, note that since $\|\widehat{\mathbb{Q}}^N_{t-1} - \mathbb{Q}_{t-1}\|_{\mathrm{E}} \to 0$ as $N \to +\infty$ (Theorem 1), one may expect that

$$\|\mathcal{M}^h_{t,\widehat{\mathbb{Q}}^N_{t-1,h}} - \mathcal{M}^h_{t,\mathbb{Q}_{t-1,h}}\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty.$$

Therefore, we intuitively need smoothness assumption on this limiting Markov kernel to establish the validity of the backward pass of SQMC. However, note that the two arguments of this kernel are "projections" in $[0,1)$ through the inverse of the Hilbert curve. Consequently, it is not clear how smoothness assumptions on the Rosenblatt transformation of $\mathcal{M}_{t,\mathbb{Q}_{t-1}}$ would translate into some regularity for the Rosenblatt transformation of $\mathcal{M}^h_{t,\mathbb{Q}_{t-1,h}}$. As shown below, a consistency result for QMC forward-backward algorithm can be established under a Hölder assumption on the CDF of $\mathcal{M}_{t,\mathbb{Q}_{t-1}}$.

To establish the consistency of Algorithm 3 we proceed in two steps. First, we consider a modified backward pass which amounts to sampling from a continuous distribution. Working with a continuous distribution allows us to focus on the technical difficulties specific to the backward step we just mentioned without being distracted by complicated discontinuity issues. Then, the result obtained for this continuous backward pass is used to deduce sufficient conditions for the consistency of Algorithm 3. If this approach in two steps greatly facilitates the analysis, the resulting conditions for the validity of QMC forward-backward smoothing have the drawback to impose that the Markov kernel m_t and the potential function G_t are bounded below away from zero (see Corollary 2 below).

5.2.3. A continuous backward pass

Following the discussion above, we consider now a modified backward pass, which amounts to transforming a QMC point set $\tilde{\mathbf{u}}^{1:N}$ in $[0,1)^{T+1}$ through the inverse Rosenblatt transformation of a continuous approximation $\widetilde{\mathsf{Q}}^N_{T,h_T}$ of $\widetilde{\mathbb{Q}}^N_{T,h_T}$.

To construct $\widetilde{\mathsf{Q}}^N_{T,h_T}$, let $\widehat{\mathsf{Q}}^N_{T,h}$ be the probability measure that corresponds to a continuous approximation of the CDF of $\widehat{\mathbb{Q}}^N_{T,h}$, which is strictly increasing on $[0, h(\mathbf{x}_T^{\sigma_T(N)})]$ with $F_{\widehat{\mathsf{Q}}^N_{T,h}}(h(\mathbf{x}_T^{\sigma_T(N)})) = 1$ and such that, under the assumptions of Theorems 1 and 2,

$$\|\widehat{\mathsf{Q}}^N_{T,h} - \widehat{\mathbb{Q}}^N_{T,h}\|_{\mathrm{E}} = \mathcal{O}(1).$$

Next, for $t \in 1:T$, let $K^N_{t,h} : [0,1) \to \mathcal{P}([0,1))$ be a Markov kernel such that:

1. Its CDF is continuous on $[0,1)\times[0,h(\mathbf{x}_{t-1}^{\sigma_{t-1}(N)})]$;

2. $\forall h_t\in[0,1)$, the CDF of $K_{t,h}^N(h_t,\mathrm{d}h_{t-1})$ is strictly increasing on $[0,h(\mathbf{x}_{t-1}^{\sigma_{t-1}(N)})]$ with $F_{K_{t,h}^N}(h_t,h(\mathbf{x}_{t-1}^{\sigma_{t-1}(N)}))=1$;

3. Under the assumptions of Theorems 1 and and 2,

$$\sup_{h_t\in[0,1)}\|K_{t,h}^N(h_t,\mathrm{d}h_{t-1})-\mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1,h}^N}^h(h_t,\mathrm{d}h_{t-1})\|_{\mathrm{E}}=o(1).$$

Finally, we define $\widetilde{\mathsf{Q}}_{T,h_T}^N\in\mathcal{P}([0,1)^{T+1})$ as

$$\widetilde{\mathsf{Q}}_{T,h_T}^N(\mathrm{d}h_{0:T}):=\widehat{\mathsf{Q}}_{T,h}^N(\mathrm{d}h_T)\prod_{t=1}^T K_{t,h}^N(h_t,\mathrm{d}h_{t-1})$$

which, by construction, has a Rosenblatt transformation which is continuous on $[0,1)^{T+1}$.

Remark that such a distribution $\widetilde{\mathsf{Q}}_{T,h_T}^N$ indeed exists. For instance, under the assumptions of Theorems 1 and 2, one can take for $\widehat{\mathsf{Q}}_{T,h}^N$ the probability distribution that corresponds to a piecewise linear approximation of the CDF of $\widehat{\mathbb{Q}}_{T,h}^N$ and, similarly, for $h_t\in[0,1)$, one can construct $K_{t,h}^N(h_t,\mathrm{d}h_{t-1})$ from a piecewise linear approximation of the CDF of $\mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1,h}^N}^h(h_t,\mathrm{d}h_{t-1})$.

For this modified backward step we obtain the following consistency result:

Theorem 6. *Let $(\tilde{\mathbf{u}}^{1:N})_{N\geq1}$ be a sequence of point sets in $[0,1)^{T+1}$ such that $D(\tilde{\mathbf{u}}^{1:N})\to0$ as $N\to+\infty$. For $n\in1:N$, let $\check{h}_{0:T}^n=F_{\widetilde{\mathsf{Q}}_{T,h_T}^N}^{-1}(\tilde{\mathbf{u}}^n)$ where $\widetilde{\mathsf{Q}}_{T,h_T}^N$ is as above. Suppose that the Assumptions of Theorem 1 and Assumptions H1-H2 of Theorem 2 hold and that, viewed as a function of $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$, $F_{\mathcal{M}_{t,\mathbb{Q}_{t-1}}}^{cdf}(\mathbf{x}_t,\mathbf{x}_{t-1})$, the CDF of $\mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathbf{x}_t,\mathrm{d}\mathbf{x}_{t-1})$, is Hölder continuous for all $t\in1:T$. Let $\check{\mathbf{x}}_{0:T}^n=H_T(\check{h}_{0:T}^n)$. Then,*

$$\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}\to0\quad\textit{as } N\to+\infty.$$

See Appendix B.5.2 for a proof.

5.2.4. A consistency result for SQMC forward-backward smoothing

We are now ready to provide conditions which ensure that QMC forward-backward smoothing (Algorithms 2 and 3) yields a consistent estimate of the smoothing distribution. The key idea of our consistency result (Corollary 2 below) is to show that, for a given point set $\tilde{\mathbf{u}}^{1:N}$, the point set $\tilde{\mathbf{x}}_{0:T}^{1:N}$ generated by Algorithm 3 becomes, as N increases, arbitrary close to the point set $\check{\mathbf{x}}_{0:T}^{1:N}$ obtained by the modified backward step described in the previous subsection.

Corollary 2. *Consider the set-up of the SQMC forward filtering-backward smoothing algorithm (Algorithms 2 and 3) and assume the following holds for $t\in0:T-1$:*

1. *$(\tilde{\mathbf{u}}^{1:N})_{N\geq 1}$ is a sequence of point sets in $[0,1)^{T+1}$ such that $D(\tilde{\mathbf{u}}^{1:N}) \to 0$ as $N \to +\infty$;*

2. *Assumptions of Theorem 1 and Assumptions H1-H2 of Theorem 2 hold;*

3. *$F^{cdf}_{\mathcal{M}_{t,\mathbb{Q}_{t-1}}}(\mathbf{x}_t, \mathbf{x}_{t-1})$ is Hölder continuous;*

4. *There exists a constant $\underline{c}_t > 0$ such that, for all $\mathbf{x}_{(t-1):(t+1)} \in \mathcal{X}^3$,*

$$G_t(\mathbf{x}_{t-1}, \mathbf{x}_t) G_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1}) m_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1}) \geq \underline{c}_t;$$

5. *$G_t(\mathbf{x}_{t-1}, \mathbf{x}_t) m_t(\mathbf{x}_{t-1}, \mathbf{x}_t)$ is uniformly continuous on $\mathcal{X}^2$.*

Then,

$$\|\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T}) - \widetilde{\mathbb{Q}}_T\|_{\mathrm{E}} \to 0 \quad \textit{as } N \to +\infty.$$

See Appendix B.5.3 for a proof. Recall that the result above implies that

$$\frac{1}{N} \sum_{n=1}^{N} \varphi(\tilde{\mathbf{x}}^n_t) \to \widetilde{\mathbb{Q}}_T(\varphi), \quad \text{as } N \to +\infty$$

for any bounded and continuous φ, as explained in Section 2.3.

Assumption 4 is the main assumption of this result. This strong condition is the price to pay for our study of QMC backward smoothing in two steps which, again, has the advantage to facilitate the analysis by avoiding complicated discontinuity problems. We conjecture that this assumption may be removed by using an approach similar to the proof of Theorem 4 in [17].

5.3. An alternative backward step

A drawback of Algorithm 3 is that it uses as an input a point set of $\tilde{\mathbf{u}}^{1:N}$ of dimension $(T+1)$, although T is often large in practice. It is well known that high-dimensional QMC point sets do not have good equidistribution properties, unless N is extremely large.

To address this issue, we may still use SQMC for the forward pass, but use as a backward pass Algorithm 3 with IID uniform variables as an input (i.e. input $\tilde{\mathbf{u}}^{1:N}$ is replaced by N independent uniforms). Our consistency results still apply, since $D(\tilde{\mathbf{u}}^{1:N}) \to 0$ with probability one in that case [28, page 167]. Of course, one cannot hope for a convergence rate better than $N^{-1/2}$ for such a hybrid approach, but the resulting algorithm may still perform better than standard (Monte Carlo) backward smoothing (for fixed N), while being simpler to implement than SQMC with a QMC backward pass based on a point set of dimension $T+1$.

More generally, we could take $\tilde{\mathbf{u}}^{1:N}$ to be some combination of point sets and uniform variables (e.g. the first k components of $\mathbf{u}^{1:N}$ is a point set, and the remaining components are independent uniforms), while still having $D(\tilde{\mathbf{u}}^{1:N}) \to 0$ [29]. However, we leave for further research the study of such an extension.

6. Numerical study

We consider the following multivariate stochastic volatility model (SV) proposed by [8]:

$$\begin{cases} \mathbf{y}_t = S_t^{1/2}\boldsymbol{\epsilon}_t, & t \geq 0 \\ \mathbf{x}_t = \boldsymbol{\mu} + \Phi(\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \Psi^{\frac{1}{2}}\boldsymbol{\nu}_t, & t \geq 1 \end{cases} \tag{11}$$

where $\mathbf{y}_t$ and $\mathbf{x}_t$ take values in $\mathbb{R}^d$, $S_t = \mathrm{diag}(\exp(x_{t1}), ..., \exp(x_{td}))$, Φ and Ψ are diagonal matrices and $(\boldsymbol{\epsilon}_t, \boldsymbol{\nu}_t) \sim \mathcal{N}_{2d}(\mathbf{0}_{2d}, C)$ with C a correlation matrix. (Since the state space is not $[0,1]^d$, our consistency results do not apply directly to this model.)

The parameters we use for the simulations are the same as in [8]: $\Phi = 0.9\mathbf{I}_d$, $\Psi = 0.1\mathbf{I}_d$, $\boldsymbol{\mu} = -9(1, \ldots, 1)^t$ and

$$C = \begin{pmatrix} 0.6\mathbf{1}_d + 0.4\mathbf{I}_d & \mathbf{0}_d \\ \mathbf{0}_d & 0.8\mathbf{1}_d + 0.2\mathbf{I}_d \end{pmatrix}$$

where $\mathbf{I}_d$, $\mathbf{0}_d$ and $\mathbf{1}_d$, are respectively the identity, all-zeros, and all-ones $d \times d$ matrices. The prior distribution for $\mathbf{x}_0$ is the stationary distribution of the process $(\mathbf{x}_t)_{t\geq 0}$. We take $d = 2$ and $T = 399$ (i.e. 400 observations).

We report results (a) for QMC full backward smoothing (Algorithm 2 for the forward pass, then Algorithm 3 for the backward pass), and (b) for marginal backward smoothing (as described in Section 5.1). In Figure 2 we first illustrate our L_2-convergence result for full backward smoothing (Theorem 5), by showing the evolution of the mean square error (MSE) as a function of N and for the estimation of the smoothing expectation $\mathbb{E}[x_{1t}|\mathbf{y}_{0:50}]$, with $t \in \{1, 10, 20, 30, 40\}$. Then, the different algorithms are compared with their Monte Carlo counterpart using the gain factors for the estimation of the smoothing expectation $\mathbb{E}[x_{1t}|\mathbf{y}_{0:T}]$, $t \in 0 : T$, which we define as the Monte Carlo (MSE) over the quasi-Monte Carlo MSE. Results for component x_{2t} of $\mathbf{x}_t$ are mostly similar (by symmetry) and thus are not reported.

The implementation of QMC and Monte Carlo algorithms are as in [17]. In SQMC, prior to the Hilbert sort step, the particles are mapped into $[0,1)^d$ using a component-wise (rescaled) logistic transform (i.e. each component is rescaled to have mean 0, variance 1, and then is applied the transform $x \to 1/(1+e^{-x})$). For SMC, systematic resampling [7] is used, and random variables are generated using standard methods (i.e. not using the inverse Rosenblatt transformation). The complete C/C++ code is available on-line at `https://bitbucket.org/mgerber/sqmc`.

Figure 3 plots the gain factors at each time step, for either $N = 2^8$ (left), or $N = 2^{10}$ (right). We observe that gain factors tend to increase with N (as expected) and that they are above one most of the time. They are not very high for full backward smoothing; but note that even a marginal improvement in terms of gain factor may translate in high CPU time savings, given that these algorithms have complexity $\mathcal{O}(N^2)$; i.e. a gain factor of 3 means that SMC would need 3 times more particles, and therefore 9 times more CPU time, to reach the same accuracy as SQMC. Notice also that gain factors are higher for marginal smoothing.

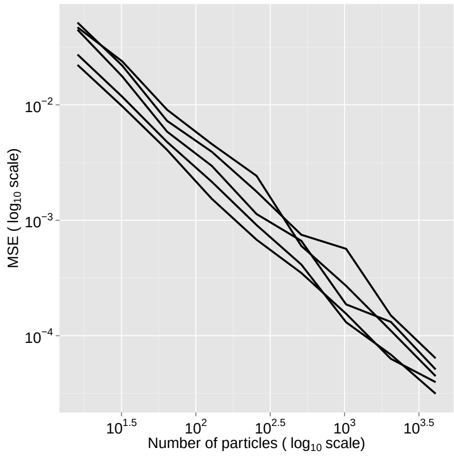

Figure 2: Full backward smoothing (Algorithm 3) of the bivariate SV model (11). The graph gives the MSE (from 100 replications) for $\mathbb{E}[x_{t1}|\mathbf{y}_{0:49}]$ as a function of N and for $t \in \{1, 10, 20, 30, 40\}$.

Finally, we compare Algorithm 3 (full backward smoothing) with the hybrid strategy described at the end of Section 5.3: i.e. a SQMC forward pass (Algorithm 2) followed by a Monte Carlo backward pass. Again, this is for $N = 2^8$ (left) and $N = 2^{10}$ (right). Interestingly, the hybrid strategy (slightly) dominates at most time steps (excepts those such that $T - t$ is small). As already discussed, the likely reason for this phenomenon is that the backward pass of Algorithm 3 is based on a point set of dimension T, which is too large to have good equidistribution properties (for reasonable values of N), and therefore to bring much improvement over plain Monte Carlo. Thus, for large T, one may as well use this hybrid strategy to perform full smoothing.

7. Conclusion

The estimation of the smoothing distribution $p(\mathbf{x}_{0:T}|\mathbf{y}_{0:T})$ is a challenging task for QMC methods because it is typically a high dimensional problem. On the other hand, due to the $\mathcal{O}(N^2)$ complexity of most smoothing algorithms, small gains in term of mean square errors translate into important savings in term of running times to reach the same level of error. In this work we provide asymptotic results for some QMC smoothing strategies, namely forward smoothing, and two variants of forward-backward smoothing. In a simulation study we show that the QMC forward-backward smoothing algorithm outperforms its Monte Carlo counterpart despite of the high dimensional nature of the problem. Also, if one is interested in the estimation of the marginal smoothing distributions, more important gains may be obtained.

The set of smoothing strategies discussed in this work is obviously not exhaustive. For instance, we have not discussed two-filter smoothing [5], or its $\mathcal{O}(N)$ variant proposed by [16]. In fact, our analysis can be easily applied to derive a QMC version of these

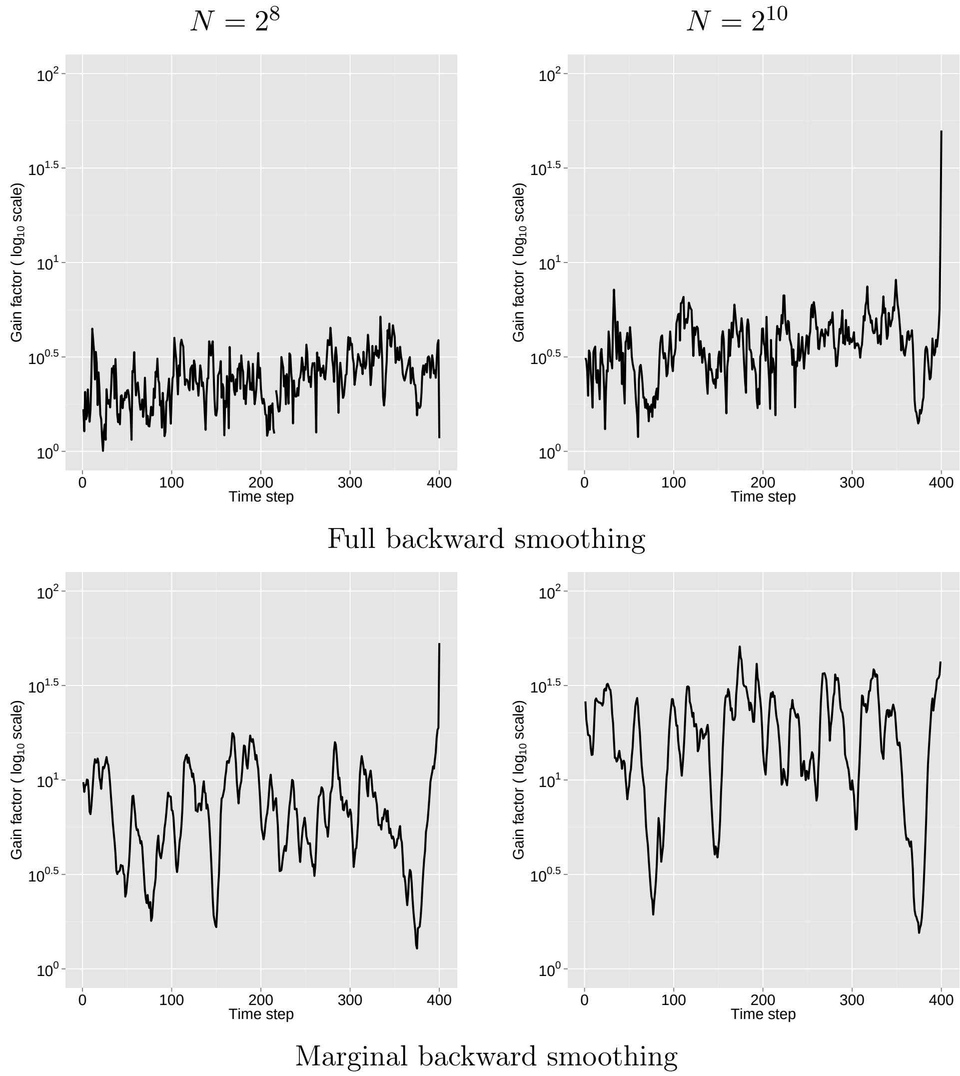

Figure 3: Smoothing of the bivariate SV model (11) for $N = 2^8$ and $N = 2^{10}$ particles. The graphs give the gain factor (MSE ratio, from 100 replications) for comparing SQMC with SMC, and for $\mathbb{E}[x_{t1}|\mathbf{y}_{0:T}]$ as a function of t. The top line is for full backward smoothing (Algorithm 3), the bottom line is for marginal backward smoothing.

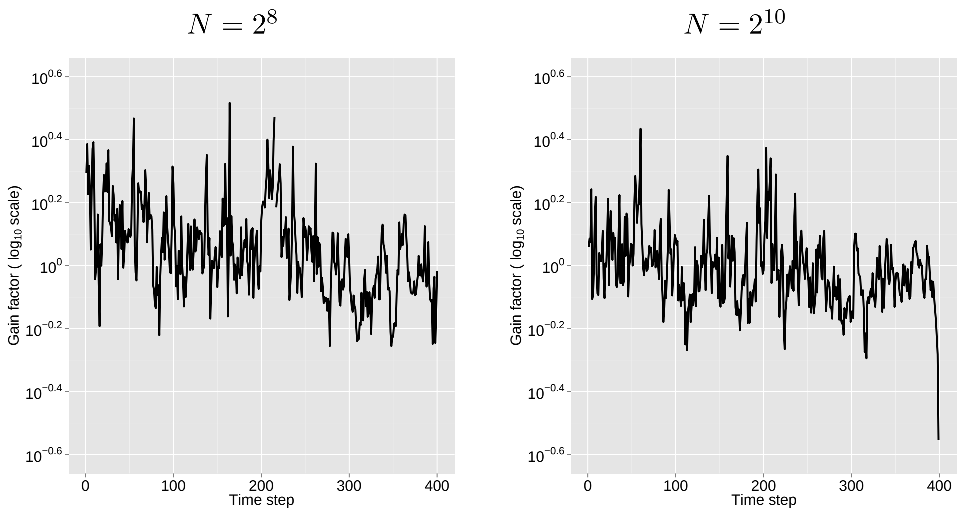

Figure 4: Smoothing of the bivariate SV model (11) for $N = 2^8$ and $N = 2^{10}$ particles. The graphs give the gain factor (MSE ratio, for 100 replications) of the hybrid backward pass (Algorithm 3 with IID input) relative to the QMC backward pass (Algorithm 3 with a QMC point set as input), for the estimation of $\mathbb{E}[x_{t1}|\mathbf{y}_{0:T}]$ as a function of t.

algorithms and to provide conditions for their validity. Another interesting smoothing algorithm is proposed in [12], where the backward pass is an accept-reject procedure, leading to a $\mathcal{O}(N)$ complexity. Yet another smoothing strategy is the particle Gibbs sampler proposed by [3] which generates a Markov chain having the smoothing distribution as stationary distribution. For these last two methods, the usefulness and the validity of replacing pseudo-random numbers by QMC point sets remain open questions.

Acknowledgements

We thank Arnaud Doucet, Art B. Owen, Florian Pelgrin and two anonymous referees for useful comments. The second author is partially supported by a grant from the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (ANR-11-LABEX-0047).

References

[1] Aistleitner, C. and Dick, J. (2013). Low-discrepancy point sets for non-uniform measures. *ArXiv preprint arXiv:1308.5049.*

[2] Aistleitner, C. and Dick, J. (2014). Functions of bounded variation, signed measures, and a general Koksma-Hlawka inequality. *ArXiv preprint arXiv:1406.0230.*

[3] Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342.

[4] Briers, M., Doucet, A., and Maskell, S. (2010). Smoothing algorithms for state–space models. *Ann. of the Inst. of Stat. Math.*, 62(1):61–89.

[5] Briers, M., Doucet, A., and Singh, S. S. (2005). Sequential auxiliary particle belief propagation. In *Proc. 8th International Conference on Information Fusion*, volume 1.

[6] Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models.* Springer-Verlag, New York.

[7] Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proc. Radar, Sonar Navigation*, 146(1):2–7.

[8] Chan, D., Kohn, R., and Kirby, C. (2006). Multivariate stochastic volatility models with correlated errors. *Econometr. Rev.*, 25(2-3):245–274.

[9] Del Moral, P. (2004). *Feynman-Kac formulae. Genealogical and interacting particle systems with applications.* Probability and its Applications. Springer Verlag, New York.

[10] Del Moral, P., Doucet, A., and Singh, S. S. (2010). A backward particle interpretation of Feynman-Kac formulae. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):947–975.

[11] Dick, J. and Pillichshammer, F. (2010). *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration.* Cambridge University Press.

[12] Douc, R., Garivier, A., Moulines, E., and Olsson, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *The Annals of Applied Probability*, 21(6):2109–2145.

[13] Doucet, A., de Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, New York.

[14] Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.*, 10(3):197–208.

[15] El Haddad, R., Lécot, C., and L'Ecuyer, P. (2008). Quasi-Monte Carlo simulation of discrete-time Markov chains on multidimensional state spaces. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 413–429. Springer.

[16] Fearnhead, P., Wyncoll, D., and Tawn, J. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447.

[17] Gerber, M. and Chopin, N. (2015). Sequential Quasi-Monte Carlo. *J. R. Statist. Soc. B*, 77(3):509–579.

[18] Godsill, S. J., Doucet, A., and West, M. (2004). Monte Carlo smoothing for nonlinear times series. *J. Amer. Statist. Assoc.*, 99(465):156–168.

[19] Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Comm., Radar, Signal Proc.*, 140(2):107–113.

[20] Hamilton, C. H. and Rau-Chaplin, A. (2008). Compact Hilbert indices for multi-dimensional data. In *Proceedings of the First International Conference on Complex, Intelligent and Software Intensive Systems*.

[21] He, Z. and Owen, A. B. (2014). Extensible grids: uniform sampling on a space-flling curve. *ArXiv preprint arXiv:1406.4549*.

[22] Hlawka, E. and Mück, R. (1972). Uber eine transformation von gleichverteilten folgen II. *Computing*, 9:127–138.

[23] L'Ecuyer, P., Lécot, C., and L'Archevêque-Gaudet, A. (2009). On array-RQMC for Markov Chains: Mapping alternatives and convergence rates. In *Monte Carlo and quasi-Monte Carlo methods 2008*, pages 485–500. Springer Berlin Heidelberg.

[24] L'Ecuyer, P., Lécot, C., and Tuffin, B. (2006). A randomized quasi-Monte Carlo simulation method for Markov chain. In *Monte-Carlo and quasi Monte-Carlo methods 2004*, pages 331–342. Springer Berlin Heidelberg.

[25] L'Ecuyer, P., Lécot, C., and Tuffin, B. (2008). A randomized quasi-monte carlo simulation method for markov chains. *Operations Research*, 56(4):958–975.

[26] Lemieux, C. (2009). *Monte Carlo and Quasi-Monte Carlo Sampling (Springer Series in Statistics)*. Springer.

[27] Leobacher, G. and Pillichshammer, F. (2014). *Introduction to quasi-Monte Carlo integration and applications*. Springer.

[28] Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. CBMS-NSF Regional conference series in applied mathematics.

[29] Ökten, G., Tuffin, B., and Burago, V. (2006). A central limit theorem and improved error bounds for hybrid-Monte Carlo sequence with applications in computational finance. *J. Complexity*, 22(4):435–458.

[30] Owen, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Lecture Notes in Statististics*, volume 106, pages 299–317. Springer, New York.

[31] Owen, A. B. (1997). Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.*, 34(5):1884–1910.

[32] Owen, A. B. (1998). Scrambling Sobol' and Niederreiter-Xing points. *J. Complexity*, 14(4):466–489.

[33] Poyiadjis, G., Doucet, A., and Singh, S. S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98:65–80.

[34] Sagan, H. (1994). *Space-Filling curves.* Springer-Verlag.

[35] Schretter, C., He, Z., Gerber, M., and Niederreiter, H. (2015). Van der corput and golden ratio sequences along the hilbert space-filling curve. In *Monte Carlo and Quasi-Monte Carlo Methods 2014 (to appear)*. Springer-Verlag.

[36] Van der Vaart, A. W. (2007). *Asymptotic Statistics.* Cambrige series in statistical and probabilistic mathematics.

[37] Wächter, C. and Keller, A. (2008). Efficient simultaneous simulation of Markov chains. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 669–684. Springer.

A. Main properties of the Hilbert curve

Function H is obtained as the limit of a certain sequence (H_m) of functions $H_m : [0,1] \to [0,1]^d$ as $m \to \infty$. The proofs of the results presented in this work are based on the following technical properties of H and H_m. For $m \geq 0$, let $\mathcal{I}_m^d = \{I_m^d(k)\}_{k=0}^{2^{md}-1}$ be the collection of consecutive closed intervals in $[0,1]$ of equal size 2^{-md} and such that $\cup\mathcal{I}_m^d = [0,1]$. For $k \geq 0$, $S_m^d(k) = H_m(I_m^d(k))$ belongs to $\mathcal{S}_m^d$, the set of the 2^{md} closed hypercubes of volume 2^{-md} that covers $[0,1]^d$, $\cup\mathcal{S}_m^d = [0,1]^d$; $S_m^d(k)$ and $S_m^d(k+1)$ are adjacent, i.e. have at least one edge in common (*adjacency property*). If we split $I_m^d(k)$ into the 2^d successive closed intervals $I_{m+1}^d(k_i)$, $k_i = 2^d k + i$ and $i \in 0 : 2^d - 1$, then the $S_{m+1}^d(k_i)$'s are simply the splitting of $S_m^d(k)$ into 2^d closed hypercubes of volume $2^{-d(m+1)}$ (*nesting property*). Finally, the limit H of H_m has the *bi-measure property*: $\lambda_1(A) = \lambda_d(H(A))$, for any measurable set $A \subset [0,1]$, and satisfies the *Hölder condition* $\|H(x_1) - H(x_2)\| \leq C_H |x_1 - x_2|^{1/d}$ for all x_1 and x_2 in $[0,1]$. For more background on space-filling curves, see [34].

B. Proofs

B.1. Backward decomposition: Proof of Theorem 2

Lemma 2 of [17] is central for the proof of this result and is reproduced here for sake of clarity.

Lemma 1. *Let $(\pi^N)_{N\geq 1}$ be a sequence of probability measures on $[0,1)^{d_1}$ such that $\|\pi^N - \pi\|_{\mathrm{E}} \to 0$ as $N \to +\infty$ for some $\pi \in \mathcal{P}([0,1)^{d_1})$, and let K a kernel $[0,1)^{d_1} \to \mathcal{P}([0,1)^{d_2}))$ such that $F_K(\mathbf{x}_1, \mathbf{x}_2)$ is Hölder continuous with its i-th component strictly increasing in x_{2i}, $i \in 1 : d_2$. Then*

$$\|\pi^N \otimes K - \pi \otimes K\|_{\mathrm{E}} = o(1).$$

From Theorem 1, we know that (for $t \geq 1$)

$$\|\mathcal{S}(P_{t,h}^N) - \mathbb{Q}_{t-1,h} \otimes m_{t,h}\|_{\mathrm{E}} = o(1) \qquad \text{for } P_{t,h}^N = \left(h(\hat{\mathbf{x}}_{t-1}^{1:N}), \mathbf{x}_t^{1:N}\right).$$

To establish (6), we fix $\mathbf{x}_{t+1}$, and recognise $\mathcal{M}_{t+1,\mathbb{Q}_t}$ as the marginal distribution of $\mathbf{x}_t$, relative to joint distribution

$$\frac{\tilde{G}_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1}) G_{t,h}(h_{t-1}, \mathbf{x}_t)}{\mathbb{Q}_{t-1,h} \otimes m_{t,h}(G_{t,h})} \times \mathbb{Q}_{t-1,h} \otimes m_{t,h}\left(\mathrm{d}(h_{t-1}, \mathbf{x}_t)\right) \tag{12}$$

with $G_{t,h}(h_{t-1}, \mathbf{x}_t) = G_t(H(h_{t-1}, \mathbf{x}_t))$. This is a change of measure applied to $\mathbb{Q}_{t-1,h} \otimes m_{t,h}$. Similarly, $\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}$ is the marginal of a joint distribution obtained by the same change of measure, but applied to $\mathcal{S}(P_{t,h}^N)$.

Thus, we may apply Theorem 1 of [17], and deduce that (again for a fixed $\mathbf{x}_{t+1}$):

$$\|\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}(\mathbf{x}_{t+1}, \mathrm{d}\mathbf{x}_t) - \mathcal{M}_{t+1,\mathbb{Q}_t}(\mathbf{x}_{t+1}, \mathrm{d}\mathbf{x}_t)\|_{\mathrm{E}} = o(1).$$

To see that the $o(1)$ term in the above expression does not depend on $\mathbf{x}_{t+1}$, note that in (12), the dominating measure does not depends on $\mathbf{x}_{t+1}$, and the density with respect to this dominating measure is bounded uniformly with respect to $\mathbf{x}_{t+1}$, and therefore the results follows from the computations in the proof of [17, Theorem 1]. This shows (6) for $t \geq 1$. For $t = 0$ replace $\mathbb{Q}_{t-1,h} \otimes m_{t,h}$ by $m_{0,h}$ in the above argument.

Let us now prove the second part of the theorem. As a preliminary result to establish (7) we show that, for all $t \geq 0$,

$$\|\widehat{\mathbb{Q}}_{t+1}^N \otimes \mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N} - \mathbb{Q}_{t+1} \otimes \mathcal{M}_{t+1,\mathbb{Q}_t}\|_{\mathrm{E}} = o(1). \tag{13}$$

Let B_t and B_{t+1} be two sets in $\mathcal{B}_{[0,1)^d}$ and note $B_{t:t+1} = B_t \times B_{t+1}$ to simplify the notations. Then,

$$\begin{aligned}
&\left|\widehat{\mathbb{Q}}_{t+1}^N \otimes \mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}(B_{t:t+1}) - \mathbb{Q}_{t+1} \otimes \mathcal{M}_{t+1,\mathbb{Q}_t}(B_{t:t+1})\right| \\
&= \left|\int_{B_{t+1}} \lambda_d\left(F_{\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}}(\mathbf{x}_{t+1}, B_t)\right) \widehat{\mathbb{Q}}_{t+1}^N(\mathrm{d}\mathbf{x}_{t+1}) - \lambda_d\left(F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, B_t)\right) \mathbb{Q}_{t+1}(\mathrm{d}\mathbf{x}_{t+1})\right| \\
&\leq \left|\int_{B_{t+1}} \lambda_d\left(F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, B_t)\right) \left(\widehat{\mathbb{Q}}_{t+1}^N - \mathbb{Q}_{t+1}\right)(\mathrm{d}\mathbf{x}_{t+1})\right| \\
&+ \left|\int_{B_{t+1}} \widehat{\mathbb{Q}}_{t+1}^N(\mathrm{d}\mathbf{x}_{t+1}) \left[\lambda_d\left(F_{\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}}(\mathbf{x}_{t+1}, B_t)\right) - \lambda_d\left(F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, B_t)\right)\right]\right|.
\end{aligned}$$

By assumption, $F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, \mathbf{x}_t)$ is Hölder continuous. Since $\|\widehat{\mathbb{Q}}_{t+1}^N - \mathbb{Q}_{t+1}\|_{\mathrm{E}} = o(1)$ by Theorem 1, Lemma 1 therefore implies

$$\sup_{B_{t:t+1} \in \mathcal{B}_{[0,1)^d}^2} \left|\int_{B_{t+1}} \lambda_d\left(F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, B_t)\right) \left(\widehat{\mathbb{Q}}_{t+1}^N - \mathbb{Q}_{t+1}\right)(\mathrm{d}\mathbf{x}_{t+1})\right| = o(1).$$

In addition,

$$
\begin{aligned}
&\left|\int_{B_{t+1}} \widehat{\mathbb{Q}}_{t+1}^N(\mathrm{d}\mathbf{x}_{t+1}) \left[\lambda_d\left(F_{\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}}(\mathbf{x}_{t+1}, B_t)\right) - \lambda_d\left(F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, B_t)\right)\right]\right| \\
&\leq \int_{B_{t+1}} \widehat{\mathbb{Q}}_{t+1}^N(\mathrm{d}\mathbf{x}_{t+1}) \sup_{B_t\in\mathcal{B}_{[0,1)^d}} \left|\lambda_d\left(F_{\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}}(\mathbf{x}_{t+1}, B_t)\right) - \lambda_d\left(F_{\mathcal{M}_{t+1,\mathbb{Q}_t}}(\mathbf{x}_{t+1}, B_t)\right)\right| \\
&\leq \int_{B_{t+1}} \widehat{\mathbb{Q}}_{t+1}^N(\mathrm{d}\mathbf{x}_{t+1}) \sup_{\mathbf{x}_{t+1}\in[0,1)^d} \|\mathcal{M}_{t+1,\widehat{\mathbb{Q}}_t^N}(\mathbf{x}_{t+1}, \mathrm{d}\mathbf{x}_t) - \mathcal{M}_{t+1,\mathbb{Q}_t}(\mathbf{x}_{t+1}, \mathrm{d}\mathbf{x}_t)\|_{\mathrm{E}} \\
&= \mathcal{o}(1)
\end{aligned}
$$

using (6). This complete the proof of (13).

We are now ready to prove the second statement of the theorem. Note that (7) is true for $t = 1$ by (13). Let $t > 1$ and $B_{0:t} \in \mathcal{B}_{[0,1)^d}^{t+1}$. Then,

$$
\begin{aligned}
&\left|\int_{B_{0:t}} \left(\widetilde{\mathbb{Q}}_t^N - \widetilde{\mathbb{Q}}_t\right)(\mathrm{d}\mathbf{x}_{0:t})\right| = \left|\int_{B_{0:t}} \left(\widehat{\mathbb{Q}}_t^N \otimes \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathrm{d}\mathbf{x}_{t-1:t}) \prod_{s=1}^{t-1} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right.\right. \\
&\left.\left. - \mathbb{Q}_t \otimes \mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathrm{d}\mathbf{x}_{t-1:t}) \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right)\right| \\
&\leq \left|\int_{B_{t-1:t}} \left[\int_{B_{0:t-2}} \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right] \left(\widehat{\mathbb{Q}}_t^N \otimes \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N} - \mathbb{Q}_t \otimes \mathcal{M}_{t,\mathbb{Q}_{t-1}}\right)(\mathrm{d}\mathbf{x}_{t-1:t})\right| \\
&+ \left|\int_{B_{t-1:t}} \widehat{\mathbb{Q}}_t^N \otimes \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathrm{d}\mathbf{x}_{t-1:t}) \left(\int_{B_{0:t-2}} \prod_{s=1}^{t-1} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1}) - \int_{B_{0:t-2}} \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right)\right|.
\end{aligned}
$$

The first term after the inequality sign can be rewritten as

$$
\left|\int_{B_{t-1:t}} \lambda_{(t-1)d}\left(F_{\otimes_{s=1}^{t-1}\mathcal{M}_{s,\mathbb{Q}_{s-1}}}(\mathbf{x}_{t-1}, B_{0:t-2})\right) \left(\widehat{\mathbb{Q}}_t^N \otimes \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N} - \mathbb{Q}_t \otimes \mathcal{M}_{t,\mathbb{Q}_{t-1}}\right)(\mathrm{d}\mathbf{x}_{t-1:t})\right|.
$$

The supremum of this quantity over $B_{0:t} \in \mathcal{B}_{[0,1)^d}^{t+1}$ is $\mathcal{o}(1)$ using (13), the fact that $F_{\otimes_{s=1}^{t-1}\mathcal{M}_{s,\mathbb{Q}_{s-1}}}$ is Hölder continuous (because $F_{\mathcal{M}_{s,\mathbb{Q}_{s-1}}}$ is Hölder continuous for all s) and Lemma 1.

To control the second term we first prove by induction that, for any $t > 1$,

$$
\sup_{B_{0:t-2}\in\mathcal{B}_{[0,1)^d}^{t-1}} \left|\int_{B_{0:t-2}} \prod_{s=1}^{t-1} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1}) - \int_{B_{0:t-2}} \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right| = \mathcal{o}(1) \tag{14}
$$

uniformly on $\mathbf{x}_{t-1}$. By (6) this result is true for $t = 2$. Assume that (14) holds for $t > 2$.

Then

$$\left|\int_{B_{0:t-1}} \prod_{s=1}^{t} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1}) - \int_{B_{0:t-1}} \prod_{s=1}^{t} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right|$$

$$= \left|\int_{B_{0:t-1}} \left[\mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) \prod_{s=1}^{t-1} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right.\right.$$

$$\left.\left. - \ \mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right]\right|$$

$$\leq \left|\int_{B_{t-1}} \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) \int_{B_{0:t-2}} \left(\prod_{s=1}^{t-1} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1}) - \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right)\right|$$

$$+ \left|\int_{B_{t-1}} \lambda_{(t-1)d}\left(F_{\otimes_{s=1}^{t-1}\mathcal{M}_{s,\mathbb{Q}_{s-1}}}(\mathbf{x}_{t-1}, B_{0:t-2})\right)\left(\mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1}) - \mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1})\right)\right|$$

where we saw above that second term on the right side of the inequality sign is $o(1)$ uniformly on $\mathbf{x}_t$ while the first term is bounded by

$$\int_{[0,1)^d} \mathcal{M}_{t,\widehat{\mathbb{Q}}_{t-1}^N}(\mathbf{x}_t, \mathrm{d}\mathbf{x}_{t-1})$$

$$\times \sup_{B_{0:t-2}\in\mathcal{B}_{[0,1)^d}^{t-1}} \left|\int_{B_{0:t-2}} \left(\prod_{s=1}^{t-1} \mathcal{M}_{s,\widehat{\mathbb{Q}}_{s-1}^N}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1}) - \prod_{s=1}^{t-1} \mathcal{M}_{s,\mathbb{Q}_{s-1}}(\mathbf{x}_s, \mathrm{d}\mathbf{x}_{s-1})\right)\right|$$

where, by the inductive hypothesis, the second factor is $o(1)$ uniformly on $\mathbf{x}_{t-1} \in [0,1)^d$. This shows that (14) is true at time $t+1$ and therefore the proof of the theorem is complete.

B.2. Generalization of [22]: Proof of Theorem 3

The proof of this result is an adaptation of the proof of [22, “Satz 2”].

In what follows, we use the shorthand $\alpha_N(B) = \mathcal{S}(\mathbf{u}^{1:N})(B) = N^{-1}\sum_{n=1}^{N} \mathbb{1}_B(u^n)$ for any set $B \subset [0,1)^d$. One has

$$\|\mathcal{S}(\mathbf{x}^{1:N}) - \pi\|_{\mathrm{E}} = \sup_{B\in\mathcal{B}_{[0,1)^d}} \left|\alpha_N\left(F_\pi(B)\right) - \lambda_d\left(F_\pi(B)\right)\right|.$$

Let $\beta = \lceil \kappa^{-1} \rceil$, $\tilde{d} = \sum_{i=0}^{d-1} \beta^i$, L an arbitrary integer, and $\mathcal{P}$ be the partition of $[0,1)^d$ in $L^{\tilde{d}}$ congruent hyperrectangles W of size $L^{-\beta^{d-1}} \times L^{-\beta^{d-2}} \times ... \times L^{-1}$. Let $B \in \mathcal{B}_{[0,1)^d}$, $\mathcal{U}_1$ the set of the elements of $\mathcal{P}$ that are strictly in $F_\pi(B)$, $\mathcal{U}_2$ the set of elements $W \in \mathcal{P}$ such that $W \cap \partial(F_\pi(B)) \neq \varnothing$, $U_1 = \cup\,\mathcal{U}_1$, $U_2 = \cup\,\mathcal{U}_2$, and $U_1' = F_\pi(B) \setminus U_1$ so that (22, “Satz 2” or 17, Theorem 4)

$$\left|\alpha_N\left(F_\pi(B)\right) - \lambda_d\left(F_\pi(B)\right)\right| \leq \left|\alpha_N(U_1) - \lambda_d(U_1)\right| + \#\mathcal{U}_2\left\{D(\mathbf{u}^{1:N}) + L^{-\tilde{d}}\right\}$$

where, under the assumption of the theorem, $|\alpha_N(U_1)-\lambda_d(U_1)|\leq L^{\tilde{d}-1}D(\mathbf{u}^{1:N})$ [see 22].

To bound $\#\mathcal{U}_2$, we first construct a partition $\mathcal{P}'$ of $[0,1)^d$ into hyperrectangles W' of size $L'^{-\beta^{d-1}}\times ...\times L'^{-1}$ such that, for all points $\mathbf{x}$ and $\mathbf{x}'$ in W', we have

$$|F_i(x_{1:i-1},x_i)-F_i(x'_{1:i-1},x'_i)|\leq L^{-\beta^{d-i}},\quad i=1,...,d \tag{15}$$

where $F_i(x_{1:i-1},x_i)$ denotes the i-th component of $F_\pi(\mathbf{x})$ (with $F_i(x_{1:i-1},x_i)=F_1(x_1)$ when $i=1$). To that effect, let $i\in 2:d$ and note that

$$\begin{aligned}|F_i(x_{1:i-1},x_i)-F_i(x'_{1:i-1},x'_i)|\leq |F_i(x_{1:i-1},x_i)-F_i(x_{1:i-1},x'_i)|\\ +|F_i(x_{1:i-1},x'_i)-F_i(x'_{1:i-1},x'_i)|.\end{aligned}$$

By Assumption 3, the probability measure $\pi_i(x_{1:i-1},\mathrm{d}x_i)$ admits a density $p_i(x_i|x_{1:i-1})$ with respect to the Lebesgue measure such that $\|p_i(\cdot|\cdot)\|_\infty<+\infty$. Therefore, the first term after the inequality sign is bounded by $\|p_i\|_\infty L'^{-\beta^{d-i}}$. For the second term, the Hölder property of F_π implies that

$$\begin{aligned}|F_i(x_{1:i-1},x'_i)-F_i(x'_{1:i-1},x'_i)| &\leq C_\pi(i-1)^{\kappa/2}(L'^{-\beta^{d+1-i}})^{\kappa}\\ &\leq C_\pi(i-1)^{\kappa/2}(L'^{-\beta^{d+1-i}})^{1/\beta}=C_\pi(i-1)^{\kappa/2}L'^{-\beta^{d-i}}\end{aligned}$$

with C_π the Hölder constant of F_π. For $i=1$, we simply have

$$|F_1(x_1)-F_1(x'_1)|\leq\|p_1\|_\infty L'^{-\beta^{d-1}}.$$

Condition (15) is therefore verified for L' the smallest integer such that $L'\geq\tilde{C}L$, for some $\tilde{C}>0$.

Remark now that $\partial(F_\pi(B))=F_\pi(\partial(B))$ since F is a continuous function. Let $R\in\partial B$ be a $(d-1)$-dimensional face of B and $\mathcal{R}$ be the set of hyper-rectangles $W'\in\mathcal{P}'$ such that $R\cap W'\neq\varnothing$. Note that $\#\mathcal{R}\leq L'^{\tilde{d}-1}\leq(\lfloor\tilde{C}L\rfloor+1)^{\tilde{d}-1}$. For each $W'\in\mathcal{R}$, take a point $\mathbf{r}^{W'}\in R\cap W'$ and define

$$\tilde{\mathbf{r}}^{W'}=F_\pi(\mathbf{r}^{W'})\in F_\pi(R).$$

Let $\tilde{\mathcal{R}}$ be the collection of hyper-rectangles $\tilde{W}$ of size $2L^{-\beta^{d-1}}\times ...\times 2L^{-1}$ (assuming L is even) and having point $\tilde{\mathbf{r}}^{W'}$, $W'\in\mathcal{R}$, as a middle point.

For an arbitrary $\mathbf{u}\in F_\pi(R)$, let $\mathbf{x}=F_\pi^{-1}(\mathbf{u})\in R$. Hence, $\mathbf{x}$ is in one hyperrectangle $W'\in\mathcal{R}$ so that using (15)

$$|u_i-\tilde{r}_i^{W'}|=|F_i(x_{1:i-1},x_i)-F_i(r^{W'}_{1:i-1},r_i^{W'})|\leq L^{-\beta^{d-i}},\quad i=1,\dots,d.$$

This shows that $\mathbf{u}$ belongs to the hyperrectangle $\tilde{W}\in\tilde{\mathcal{R}}$ with centre $\tilde{\mathbf{r}}^{W'}$ so that $F_\pi(R)$ is covered by at most $\#\tilde{\mathcal{R}}=\#\mathcal{R}\leq(\lfloor\tilde{C}L\rfloor+1)^{\tilde{d}-1}$ hyperrectangles $\tilde{W}\in\tilde{\mathcal{R}}$. To go back to the initial partition of $[0,1)^d$ with hyperrectangles in $\mathcal{P}$, remark that every hyperrectangle in $\tilde{\mathcal{R}}$ is covered by at most c_1 hyperrectangles in $\mathcal{P}$ for a constant c_1. Finally, since the set ∂B is made of the union of $2d$ $(d-1)$-dimensional faces of B, we have $\#\mathcal{U}_2 \leq c_2 L^{\tilde{d}-1}$ for a constant c_2.

Then, we may conclude the proof as follows

$$\|\mathcal{S}(\mathbf{x}^{1:N}) - \pi\|_{\mathrm{E}} \leq L^{\tilde{d}-1} D(\mathbf{u}^{1:N}) + c_2 L^{\tilde{d}-1} \left(D(\mathbf{u}^{1:N}) + L^{-\tilde{d}} \right)$$

where the optimal value of L is such that, for some $c_3 > 0$,

$$\|\mathcal{S}(\mathbf{x}^{1:N}) - \pi\|_{\mathrm{E}} \leq c_3 D(\mathbf{u}^{1:N})^{1/\tilde{d}}.$$

B.3. Consistency of forward smoothing: Proof of Proposition 1

The proof amounts to a simple adaptation of Theorem 1: by replacing Assumption 4 by Assumption 4' above, one obtains that $\|\mathcal{S}(\tilde{P}^N_{t,h^t}) - \widetilde{\mathbb{Q}}_{t-1,h^t} \otimes m_{t,h}\|_{\mathrm{E}} \to 0$ as $N \to +\infty$, where $\tilde{P}^N_{t,h^t} = \left(h^t(\hat{\mathbf{z}}^{1:N}_{t-1}), \mathbf{x}^{1:N}_t\right)$ $\widetilde{\mathbb{Q}}_{t-1,h^t}$ is the image by h^t of $\widetilde{\mathbb{Q}}_{t-1}$, and $m_{t,h}$ is defined as in Theorem 1. Therefore, by Corollary 1,

$$\|\mathcal{S}(\mathbf{z}^{1:N}_t) - \widetilde{\mathbb{Q}}_{t-1} \otimes m_t\|_{\mathrm{E}} \to 0, \quad \text{as } N \to +\infty. \tag{16}$$

In addition, since the Radon-Nikodym derivative

$$\frac{\widetilde{\mathbb{Q}}_t}{\widetilde{\mathbb{Q}}_{t-1} \otimes m_t} (\mathrm{d}(\mathbf{x}_{0:t-1}, \mathbf{x}_t)) \propto G_t(\mathbf{x}_{t-1}, \mathbf{x}_{t-1}),$$

is continuous and bounded, Theorem 1 of [17], together with (16), implies (9).

B.4. L_2-convergence: Proof of Theorem 5

To prove the result, let φ be as in the statement of the theorem and let us first prove the L_1-convergence.

We have

$$\mathbb{E}\left|\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi) - \widetilde{\mathbb{Q}}_T(\varphi)\right| \leq \mathbb{E}\left|\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi) - \widetilde{\mathbb{Q}}^N_T(\varphi)\right| + \mathbb{E}\left|\widetilde{\mathbb{Q}}^N_T(\varphi) - \widetilde{\mathbb{Q}}_T(\varphi)\right|.$$

By portmanteau lemma [36, Lemma 2.2, p.6], convergence in the sense of the extreme metric is stronger than weak convergence. Hence, the second term above goes to 0 as $N \to +\infty$ by Theorem 2 and by the dominated convergence theorem. For the first term, as each $\tilde{\mathbf{u}}^n \sim \mathcal{U}([0,1)^{T+1})$, we have, by the inverse Rosenblatt interpretation of the backward pass of SQMC,

$$\mathbb{E}\Big[\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi)\big|\mathcal{F}_T\Big] = \mathbb{E}\Big[\mathcal{S}(\tilde{h}^{1:N}_{0:T})(\varphi \circ H_T)\big|\mathcal{F}_T\Big] = \widetilde{\mathbb{Q}}^N_{T,h_T}(\varphi \circ H_T) = \widetilde{\mathbb{Q}}^N_T(\varphi)$$

with $\mathcal{F}^N_T$ the σ-algebra generated by the forward step (Algorithm 2). Therefore,

$$\mathbb{E}\left[\left|\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi) - \widetilde{\mathbb{Q}}^N_T(\varphi)\right| \big|\mathcal{F}_T\right] \leq \mathrm{Var}\Big(\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})(\varphi)\big|\mathcal{F}_T\Big)^{1/2} \tag{17}$$

where, using Assumption 2 and the fact that $\tilde{\mathbf{x}}_{0:T}^n = H_T \circ F^{-1}_{\widetilde{\mathbb{Q}}^N_{T,h_T}}(\tilde{\mathbf{u}}^n)$,

$$\mathrm{Var}\Big(\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi)|\mathcal{F}_T^N\Big) \leq Cr(N)\sigma^2_{\varphi,N} \tag{18}$$

with $\sigma^2_{\varphi,N} \leq \widetilde{\mathbb{Q}}_T^N(\varphi^2)$ and with C and $r(N)$ as in the statement of the theorem. Let $\epsilon > 0$. Then, by Assumption 1 and looking at the proof of Theorem 2, we have for N large enough and almost surely, $\widetilde{\mathbb{Q}}_T^N(\varphi^2) \leq \widetilde{\mathbb{Q}}_T(\varphi^2) + \epsilon$ so that, for N large enough,

$$\mathbb{E}\left|\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi) - \widetilde{\mathbb{Q}}_T^N(\varphi)\right| \leq \sqrt{Cr(N)\big(\widetilde{\mathbb{Q}}_T(\varphi^2)+\epsilon\big)} \tag{19}$$

showing the L_1-convergence. To prove the L_2-convergence, remark that

$$\mathbb{E}\left[\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi)|\mathcal{F}_T^N\right] = \widetilde{\mathbb{Q}}_T^N(\varphi) = \big(\widetilde{\mathbb{Q}}_T^N(\varphi) - \widetilde{\mathbb{Q}}_T(\varphi)\big) + \widetilde{\mathbb{Q}}_T(\varphi)$$

and therefore

$$\mathrm{Var}\Big(\mathbb{E}\Big[\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi)|\mathcal{F}_T^N\Big]\Big) = \mathrm{Var}\Big(\widetilde{\mathbb{Q}}_T^N(\varphi) - \widetilde{\mathbb{Q}}_T(\varphi)\Big) \leq \mathbb{E}\Big[\big(\widetilde{\mathbb{Q}}_T^N(\varphi) - \widetilde{\mathbb{Q}}_T(\varphi)\big)^2\Big]$$

where the right-hand side converges to zero as $N \to +\infty$ by the dominated convergence theorem and by Theorem 2. On conclude the prove using (17)-(19) and the fact that

$$\mathrm{Var}\Big(\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi)\Big) = \mathrm{Var}\Big(\mathbb{E}\Big[\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi)|\mathcal{F}_T^N\Big]\Big) + \mathbb{E}\Big[\mathrm{Var}\Big(\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(\varphi)|\mathcal{F}_T^N\Big)\Big].$$

B.5. Consistency of the Backward step: Proof of Theorem 6 and proof of Corollary 2

B.5.1. Preliminary computations

To prove Theorem 6 we need the following two lemmas:

Lemma 2. *Let $m \in \mathbb{N}$, $I = [0, \frac{k+1}{2^{dm}}]$, $k \in \{0, 1, ..., 2^{dm} - 2\}$ and $B = H(I)$. Then, $B = \cup_{i=1}^p B_i$ for some closed hyperrectangles $B_i \subseteq [0,1]^d$ and where $p \leq 2^d(m+1)$.*

Proof. To prove the Lemma, let $0 \leq m_1 \leq m$ be the smallest integer $\tilde{m}$ such that $I^d_{\tilde{m}}(0) \subseteq I$ and $i^*_{m_1}$ be the number of intervals in $\mathcal{I}^d_{m_1}$ included in I. Note that $i^*_{m_1} < 2^d$. Indeed, if $i^*_{m_1} \geq 2^d$ then, by the nesting property of the Hilbert curve,

$$I^d_{m_1-1}(0) \subseteq \bigcup_{k=0}^{2^d-1} I^d_{m_1}(k) \subseteq \bigcup_{k=0}^{i^*_{m_1}-1} I^d_{m_1}(k) \subseteq I$$

which is in contradiction with the definition of $i^*_{m_1}$. Define $I_2 = I \setminus \cup \mathcal{I}^I_{m_1}$ and $i^*_{m_2}$ the number of intervals in $\mathcal{I}^d_{m_2}$ included in I_2. For the same reason as above $i^*_{n_2} < 2^d$. More generally, for any $m_1 \leq m_k \leq m$, $i^*_{m_k} \leq 2^d$ meaning that the set B is made of at most $\sum_{k=m_1}^m i^*_{m_k} \leq 2^d(m+1)$ hypercubes (of side varying between 2^{-m} and 2^{-m_1}). □

Lemma 3. *Let $(\pi^N)_{N\geq 1}$ be a sequence of probability measures on $[0,1)^{(k+1)d}$ such that $\|\pi^N-\pi\|_{\mathrm{E}}\to 0$ where $\pi(\mathrm{d}\mathbf{x})=\pi(\mathbf{x})\lambda_{(k+1)d}(\mathrm{d}\mathbf{x})$ is a probability measure on $\pi^{(k+1)d}$ that admits a bounded density $\pi(\mathbf{x})$. Let π_{h_k} be the image by h_k of π. Then,*

$$\|\pi^N_{h_k}-\pi_{h_k}\|_{\mathrm{E}}\to 0,\quad \text{as } N\to+\infty.$$

The proof of this last result is omitted since it follows from the properties of Cartesian products and from straightforward modifications of the proof of [17, Theorem 3].

B.5.2. Proof of the Theorem 6

To prove the theorem first note that

$$\|\widetilde{\mathsf{Q}}^N_{T,h_T}-\widetilde{\mathbb{Q}}_{T,h_T}\|_{\mathrm{E}}=o(1).$$

Indeed, by assumption, $\|\widehat{\mathsf{Q}}^N_{T,h}-\widehat{\mathbb{Q}}^N_{T,h}\|_{\mathrm{E}}=o(1)$ and, by Theorem 1 and [17, Theorem 3], $\|\widehat{\mathbb{Q}}^N_{T,h}-\mathbb{Q}_{T,h}\|_{\mathrm{E}}=o(1)$ since $\mathbb{Q}_T$ admits a bounded density (Assumption 4 of Theorem 1). Hence, $\|\widehat{\mathsf{Q}}^N_{T,h}-\mathbb{Q}_{T,h}\|_{\mathrm{E}}=o(1)$ and thus, by Theorem 4, $\|\widehat{\mathsf{Q}}^N_{T}-\mathbb{Q}_{T}\|_{\mathrm{E}}=o(1)$, with $\widehat{\mathsf{Q}}^N_T$ the image by H of $\widehat{\mathsf{Q}}^N_{T,h}$. In addition, using the same argument, and using the fact that, for all $t\in 1:T$, $\tilde{G}_t$ is bounded (Assumption H1 of Theorem 2), we have, by Theorem 2 (first part),

$$\sup_{\mathbf{x}_t\in\mathcal{X}}\|K^N_t(\mathbf{x}_t,\mathrm{d}\mathbf{x}_{t-1})-\mathcal{M}_{t,\mathbb{Q}_{t-1}}(\mathbf{x}_t,\mathrm{d}\mathbf{x}_{t-1})\|_{\mathrm{E}}=o(1),\quad t\in 1:T$$

with $K^N_t(\mathbf{x}_t,\mathrm{d}\mathbf{x}_{t-1})$ the image by H of the probability measure $K_{t,h}(H(\mathbf{x}_t),\mathrm{d}h_{t-1})$. Consequently, by the second part of Theorem 2, $\|\widetilde{\mathsf{Q}}^N_T-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}=o(1)$ where $\widetilde{\mathsf{Q}}^N_T$ denotes the image by H_T of $\widetilde{\mathsf{Q}}^N_{T,h_T}$. Finally, under the assumptions of the theorem, $\widetilde{\mathbb{Q}}_T$ admits a bounded density (because for all t, $\tilde{G}_t$ is bounded and $\mathbb{Q}_t$ admits a bounded density) and thus, by Lemma 3, $\|\widetilde{\mathsf{Q}}^N_{T,h_T}-\widetilde{\mathbb{Q}}_{T,h_T}\|_{\mathrm{E}}=o(1)$.

To prove the theorem it therefore remains to show that

$$\|\mathcal{S}(\check{h}^{1:N}_{0:T})-\widetilde{\mathsf{Q}}^N_{T,h_T}\|_{\mathrm{E}}=o(1).\tag{20}$$

Indeed, this would yield $\|\mathcal{S}(\check{h}^{1:N}_{0:T})-\widetilde{\mathbb{Q}}_{T,h_T}\|_{\mathrm{E}}=o(1)$ and thus, by Theorem 4,

$$\|\mathcal{S}(\check{\mathbf{x}}^{1:N}_{0:T})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}=o(1)$$

as required.

To prove (20), we assume to simplify the notations that $F_{\mathcal{M}_{t,\mathbb{Q}_{t-1}}}(\mathbf{x}_t,\mathbf{x}_{t-1})$ is Lipschitz. Generalization for any Hölder exponent can be done using similar arguments as in the proof of Theorem 3.

Let $h^n_t=h(\mathbf{x}^N_t)$ where $\mathbf{x}^{1:N}_t$ are the particles obtained at the end of iteration t of Algorithm 2. We assume that, for all $t\in 0:T$, the particles are sorted according to their Hilbert index, i.e. $n<m\implies h^n_t<h^m_t$ (note that the inequality is strict by

Assumption 1 of Theorem 1). Then, using the same notation as in the proof of Theorem 3, one has

$$\|\mathcal{S}(\check{h}_{0:T}^{1:N}) - \widetilde{\mathsf{Q}}_{T,h_T}^N\|_{\mathrm{E}} = \sup_{B\in\mathcal{B}^N_{[0,1)^{T+1}}} \left|\alpha_N\left(F_{\widetilde{\mathsf{Q}}_{T,h_T}^N}(B)\right) - \lambda_{T+1}\left(F_{\widetilde{\mathsf{Q}}_{T,h_T}^N}(B)\right)\right|$$

where $\mathcal{B}^N_{[0,1)^{T+1}} = \left\{[\boldsymbol{a},\boldsymbol{b}] \subset \mathcal{B}_{[0,1)^{T+1}},\, b_i^N \leq h_i^N,\, i\in 0:T\right\}$.

The beginning of the proof follows the lines of Theorem 3, with $\beta = d$ and d replaced by $T+1$. Let $\tilde{d} = \sum_{t=0}^T d^t$ so that, for a set $B \in \mathcal{B}^N_{[0,1)^{T+1}}$,

$$\left|\alpha_N\left(F_{\widetilde{\mathsf{Q}}_{T,h_T}^N}(B)\right) - \lambda_{T+1}\left(F_{\widetilde{\mathsf{Q}}_{T,h_T}^N}(B)\right)\right| \leq L^{\tilde{d}} D(\mathbf{u}^{1:N}) + \#\mathcal{U}_2\left\{D(\mathbf{u}^{1:N}) + L^{-\tilde{d}}\right\}$$

where L and $\mathcal{U}_2$ are as in the proof of Theorem 3.

Following this latter, let $\mathcal{P}'$ be the partition of the set $[0,1)^{T+1}$ into hyperrectangles W' of size $L'^{-d^T} \times L'^{-d^{T-1}} \times \ldots \times L'^{-1}$ such that, for all $\boldsymbol{h}$ and $\boldsymbol{h}'$ in W', we have

$$|F_{\widehat{\mathsf{Q}}_{T,h}^N}(h_t) - F_{\widehat{\mathsf{Q}}_{T,h}^N}(h_1')| \leq L^{-d^T}. \tag{21}$$

and

$$\left|\tilde{F}_{i-1}^N(h_{i-1},h_i) - \tilde{F}_{i-1}^N(h_{i-1}',h_i')\right| \leq L^{-d^{T+1-i}}, \quad i \in 2:(T+1) \tag{22}$$

where, to simplify the notation, we write $\tilde{F}_{i-1}^N(\tilde{h},\cdot)$ the CDF of $K_{T-i+2,h}^N(\tilde{h}, \mathrm{d}h_{T-i+1})$.

Let us first look at condition (21). We have

$$\begin{aligned}|F_{\widehat{\mathsf{Q}}_{T,h}^N}(h_1) - F_{\widehat{\mathsf{Q}}_{T,h}^N}(h_1')| &\leq 2\|F_{\widehat{\mathbb{Q}}_{T,h}^N} - F_{\widehat{\mathsf{Q}}_{T,h}^N}\|_\infty + 2\|F_{\widehat{\mathsf{Q}}_{T,h}^N} - F_{\mathbb{Q}_{T,h}}\|_\infty + |F_{\mathbb{Q}_{T,h}}(h_1) - F_{\mathbb{Q}_{T,h}}(h_1')| \\ &\leq 2r_1(N) + 2r_2(N) + \left|F_{\mathbb{Q}_{T,h}}(h_1) - F_{\mathbb{Q}_{T,h}}(h_1')\right|\end{aligned}$$

with $r_1(N) = \|F_{\widehat{\mathsf{Q}}_{T,h}^N} - F_{\widehat{\mathbb{Q}}_{T,h}^N}\|_\infty$ and $r_2(N) = \|\widehat{\mathbb{Q}}_{T,h}^N - \mathbb{Q}_{T,h}\|_{\mathrm{E}}$; note $r_1(N) \to 0$ by the construction of $\widehat{\mathsf{Q}}_{T,h}^N$ and under the assumptions of the theorem while $r_2(N) \to 0$ by Theorem 1 and by [17, Theorem 3]

Let $L' = 2^m$ for an integer $m \geq 0$, so that h_i and h_i' are in the same interval $I^d_{d^{T-i}m}(k) \in \mathcal{I}^d_{d^{T-i}m}$, $i \in 1:(T+1)$. Then, since h_1 and h_1' are in the same interval $I^d_{d^{T-1}m}(k) \in \mathcal{I}^d_{d^{T-1}m}$,

$$\left|F_{\mathbb{Q}_{T,h}}(h_1) - F_{\mathbb{Q}_{T,h}}(h_1')\right| \leq \mathbb{Q}_{T,h}\left(I^d_{d^{T-1}m}(k)\right) = \mathbb{Q}_T\left(S^d_{d^{T-1}m}(k)\right) \leq \frac{\|p_T\|_\infty}{(L')^{d^T}}$$

as $\mathbb{Q}_T$ admits a bounded density p_T. Hence (21) is verified if

$$L' \geq L\tilde{k}_N, \quad \tilde{k}_N = \left(\frac{\|p_T\|_\infty}{(1 - L^{d^T} r_1^*(N))}\right)^{1/d^T}, \quad r_1^*(N) = 2r_1(N) + 2r_2(N),$$

which implies that we assume from now on that $L^{-d^T} \geq 2r_1^*(N)$ for N large enough.

Let us now look at (22) for a $i > 1$. To simplify the notation in what follows, let $F_{i-1}^N(\tilde{h}, \cdot)$ be the CDF of $\mathcal{M}^h_{T-i+2,\widehat{\mathbb{Q}}^N_{T-i+1,h}}(\tilde{h}, \mathrm{d}h_{T-i+1})$ and $F_{i-1}(\tilde{h}, \cdot)$ be the CDF of $\mathcal{M}^h_{T-i+2,\mathbb{Q}_{T-i+1}}(\tilde{h}, \mathrm{d}h_{T-i+1})$. Then,

$$\begin{aligned}
&\left|\tilde{F}_{i-1}^N(h_{i-1}, h_i) - \tilde{F}_{i-1}^N(h'_{i-1}, h'_i)\right| \\
&\qquad \leq 2\|\tilde{F}_{i-1}^N - F_{i-1}^N\|_\infty + 2\|F_{i-1}^N - F_{i-1}\|_\infty + |F_{i-1}(h_{i-1}, h_i) - F_{i-1}(h'_{i-1}, h'_i)| \\
&\qquad = 2r_3(N) + 2r_4(N) + \left|F_{i-1}(h_{i-1}, h_i) - F_{i-1}(h'_{i-1}, h'_i)\right|
\end{aligned}$$

with $r_3(N) = \|\tilde{F}_{i-1}^N - F_{i-1}^N\|_\infty$ and $r_4(N) = \|F_{i-1}^N - F_{i-1}\|_\infty$; note $r_3(N) \to 0$ by the construction of $K^N_{T-i+2,h}$ and under the assumptions of the theorem while $r_4(N) \to 0$ by Theorem 2 and [17, Theorem 3].

To control $\left|F_{i-1}(h_{i-1}, h_i) - F_{i-1}(h'_{i-1}, h'_i)\right|$, assume without loss of generality that $h_i \geq h'_i$ and write $\tilde{G}^h_i(h_{i-1}, h'_i) = \tilde{G}_{T-i+2}(H(h_{i-1}), H(h'_i))$ to simplify further the notation. Then

$$\begin{aligned}
|F_{i-1}(h_{i-1}, h_i) - F_{i-1}(h'_{i-1}, h'_i)| &\leq |F_{i-1}(h_{i-1}, h'_i) - F_{i-1}(h'_{i-1}, h'_i)| \\
&\quad + \left|\int_{h'_i}^{h_i} \tilde{G}^h(h_{i-1}, v)\mathbb{Q}_{T-i+1,h}(\mathrm{d}v)\right|.
\end{aligned}$$

The second term is bounded by $\|\tilde{G}_{T-i+2}\|_\infty \mathbb{Q}_{T-i+1,h}([h'_i, h_i]) \leq \|\tilde{G}_{T-i+2}\|_\infty \mathbb{Q}_{T-i+1}(W)$ where $W \in \mathcal{S}^d_{d^{T-i}m}$. Since $\mathbb{Q}_{T-i+1}$ admits a bounded density, we have, for a constant $c > 0$,

$$\|\tilde{G}_{T-i+2}\|_\infty \mathbb{Q}_{T-i+1,h}([h'_i, h_i]) \leq cL^{-d^{T+1-i}}.$$

To control the other term suppose first that $h'_i > L'^{-d^{T-i+1}}$ and let k be the largest integer such that $h'_i \geq kL'^{-d^{T-i+1}}$. Then,

$$\begin{aligned}
&|F_{i-1}(h_{i-1}, h'_i) - F_{i-1}(h'_{i-1}, h'_i)| \\
&\qquad = \left|\int_0^{h'_i} \left[\tilde{G}^h_i(h_{i-1}, v) - \tilde{G}^h_i(h'_{i-1}, v)\right] \mathbb{Q}_{T-i+1,h}(\mathrm{d}v)\right| \\
&\qquad \leq \left|\int_0^{kL'^{-d^{T-i+1}}} \left[\tilde{G}^h_i(h_{i-1}, v) - \tilde{G}^h_i(h'_{i-1}, v)\right] \mathbb{Q}_{T-i+1,h}(\mathrm{d}v)\right| \\
&\qquad + \left|\int_{kL'^{-d^{T-i+1}}}^{h'_i} \left[\tilde{G}^h_i(h_{i-1}, v) - \tilde{G}^h_i(t'_{i-1}, v)\right] \mathbb{Q}_{T-i+1,h}(\mathrm{d}v)\right|.
\end{aligned} \tag{23}$$

Then, using by Lemma 2, we have for the first term:

$$
\begin{aligned}
\Big| \int_0^{kL'^{-d^{T-i+1}}} & \Big[\tilde{G}_i^h(h_{i-1}, v) - \tilde{G}_i^h(h'_{i-1}, v)\Big] \mathbb{Q}_{T-i+1,h}(\mathrm{d}v)\Big| \\
&= \left| \sum_{j=1}^{k_i} \int_{W_j} \Big[\tilde{G}_{T-i+2}(H(h_{i-1}), \mathbf{x}) - \tilde{G}_{T-i+2}(H(h'_{i-1}), \mathbf{x})\Big] \mathbb{Q}_{T-i+1}(\mathrm{d}\mathbf{x}) \right| \\
&\leq \sum_{j=1}^{k_i} \Big\{ \Big| F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}(H(h_{i-1}), \mathbf{a_j}) - F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}(H(h'_{i-1}), \mathbf{a_j}) \Big| \\
&+ \Big| F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}(H(h_{i-1}), \mathbf{b_j}) - F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}(H(h'_{i-1}), \mathbf{b_j}) \Big| \Big\}
\end{aligned}
$$

where $W_j = [\mathbf{a}_j, \mathbf{b_j}] \subset [0,1)^d$ and where $k_i \leq 2^d(d^{T-i}m + 1)$. Let C_i be the Lipschitz constant of $F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}$. Then, for any $\mathbf{c} \in [0,1)^d$,

$$
\begin{aligned}
\Big| F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}(H(h_{i-1}), \mathbf{c}) - F^{cdf}_{\mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}}(H(h'_{i-1}), \mathbf{c}) \Big| &\leq C_i \|H(h_{i-1}) - H(h'_{i-1})\|_\infty \\
&\leq C_i L'^{-d^{T-i+1}}
\end{aligned}
$$

because $H(h_{i-1})$ and $H(h'_{i-1})$ belong to the same hypercube $W \in \mathcal{S}^d_{d^{T-i}m}$ of side $2^{-md^{T-i+1}} = L'^{-d^{T-i+1}}$.

For the second term after the inequality sign in (23), we have

$$
\begin{aligned}
&\left| \int_{kL'^{-d^{T-i+1}}}^{h'_i} \Big[\tilde{G}_i^h(h_{i-1}, v) - \tilde{G}_i^h(h'_{i-1}, v)\Big] \mathbb{Q}_{T-i+1,h}(\mathrm{d}v) \right| \\
&\leq \mathcal{M}^h_{T-i+2,\mathbb{Q}_{T-i+1,h}}(h_{i-1}, [kL'^{-d^{T-i+1}}, h'_i]) + \mathcal{M}^h_{T-i+2,\mathbb{Q}_{T-i+1,h}}(h'_{i-1}, [kL'^{-d^{T-i+1}}, h'_i]) \\
&\leq \mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}(H(h_{i-1}), W) + \mathcal{M}_{T-i+2,\mathbb{Q}_{T-i+1}}(H(h'_{i-1}), W) \\
&\leq 2\|\tilde{G}_{T-i+2}\|_\infty \|p_{T-i+1}\|_\infty 2^{-md^{T-i+1}}
\end{aligned}
$$

for a $W \in \mathcal{S}^d_{d^{T-i}m}$ and where p_{T-i+1} is the (bounded) density of $\mathbb{Q}_{T-i+1}$. This last quantity is also the bound we obtain for $h'_i < L'^{-d^{T-i+1}}$. Hence, these computations shows that

$$
|\tilde{F}_{i-1}(h_{i-1}, h_i) - \tilde{F}_{i-1}(h'_{i-1}, h'_i)| \leq c_i L'^{-d^{T-i+1}} \log(L')
$$

for a constant c_i, $i \in 2 : (T+1)$.

Condition (22) is therefore verified when (taking L' so that $\log(L') \geq 1$)

$$
\frac{L'}{\log(L')} \geq L \max_{i \in \{2,\dots,T+1\}} \left(\frac{c_i}{1 - L^{d^{T-i+1}} r_2^*(N)} \right)^{\frac{1}{d^{T-i+1}}}
$$

where $r_2^*(N) = 2r_3(N) + 2r_4(N)$. Let $\gamma \in (0,1)$ and note that for N large enough $\log L' < L'^\gamma$. Hence, for N large enough (21) and (22) are verified for L' the smallest power of 2 such that

$$L' \geq (k_N L)^{(1-\gamma)^{-1}}, \quad k_N = \max_{i \in \{1,\dots,T+1\}} \left(\frac{c_i}{1 - L^{d^{T-i+1}} r^*(N))} \right)^{\frac{1}{d^{T-i+1}}}, \quad c_1 = \|p_T\|$$

where $r^*(N) = r_1^*(N) + r_2^*(N)$. Note that we assume from now on that $L^{-d^T} \geq 2r^*(N)$.

Because the function $F_{\widetilde{\mathsf{Q}}^N_{T,h_T}}$ is continuous on $[0, h_0^N] \times \cdots \times [0, h_T^N]$, $\partial(F_{\widetilde{\mathsf{Q}}^N_{T,h_T}}(B)) = F_{\widetilde{\mathsf{Q}}^N_{T,h_T}}(\partial(B))$ and therefore we can bound $\#\mathcal{U}_2$ following the proof of Theorem 3. Using the same notations as in the proof of Theorem 3, we obtain that $\widetilde{\mathsf{Q}}^N_{T,h_T}(\partial(B))$ is covered by at most

$$(T+1)2^{\tilde{d}} k_N^{\frac{\tilde{d}-1}{1-\gamma}} L^{\frac{\tilde{d}-1}{1-\gamma}}$$

hyperrectangles in $\tilde{\mathcal{R}}$. To go back to the initial partition of $[0,1)^{T+1}$ with hyperrectangles $W \in \mathcal{P}$, remark that $L' > L$ so that every hyperrectangles in $\tilde{\mathcal{R}}$ is covered by at most c^* hyperrectangles of $\mathcal{P}$ for a constant c^*. Hence,

$$\#\mathcal{U}_2^{(1)} \leq c_N L^{\frac{\tilde{d}-1}{1-\gamma}}, \quad c_N = c^*(T+1)2^{\tilde{d}} k_N^{\frac{\tilde{d}-1}{1-\gamma}}. \tag{24}$$

We therefore have

$$\|\mathcal{S}(\check{h}_{0:T}^{1:N}) - \widetilde{\mathsf{Q}}^N_{T,h_T}\|_{\mathrm{E}} \leq L^{\tilde{d}} D(\mathbf{u}^{1:N}) + c_N L^{\frac{\tilde{d}-1}{1-\gamma}} \left(D(\mathbf{u}^{1:N}) + L^{-\tilde{d}} \right).$$

Let $\gamma \in (0, \tilde{d}^{-1})$ so that $c_d := \tilde{d} - \frac{\tilde{d}-1}{1-\gamma} > 0$. To conclude the proof as in [17, Theorem 4], let $\tilde{d}_1 = d^T$ and $\tilde{d}_2 = \sum_{t=0}^{T-1} d^t$. Thus,

$$\|\mathcal{S}(\check{h}_{0:T}^{1:N}) - \widetilde{\mathsf{Q}}^N_{T,h_T}\|_{\mathrm{E}} \leq 2L^{\tilde{d}_1+\tilde{d}_2} D(\mathbf{u}^{1:N}) + c_N L^{-c_d}$$

where the optimal value of L is such that $L = \mathcal{O}(D(\mathbf{u}^{1:N})^{-\frac{1}{c_d+\tilde{d}_1+\tilde{d}_2}})$. Then, provided that $r^*(N) D(\mathbf{u}^{1:N})^{-\frac{\tilde{d}_1}{c_d+\tilde{d}_1+\tilde{d}_2}} = \mathcal{O}(1)$, L verifies all the conditions above and, since $c_N = \mathcal{O}(1)$, we have

$$\|\mathcal{S}(\check{h}_{0:T}^{1:N}) - \widetilde{\mathsf{Q}}^N_{T,h_T}\|_{\mathrm{E}} = \mathcal{O}\left(D(\mathbf{u}^{1:N})^{\frac{1}{c_d+\tilde{d}_1+\tilde{d}_2}} \right).$$

Otherwise, if $r^*(N) D(\mathbf{u}^{1:N})^{-\frac{\tilde{d}_1}{c_d+\tilde{d}_1+\tilde{d}_2}} \to +\infty$, let $L = \mathcal{O}(r^*(N)^{-\frac{1}{\tilde{d}_1}})$. Then $c_N = \mathcal{O}(1)$ and

$$\begin{aligned}
L^{\tilde{d}_1+\tilde{d}_2} D(\mathbf{u}^{1:N}) &= \mathcal{O}(r(N))^{\frac{c_d}{\tilde{d}_1} - \frac{c_d+\tilde{d}_1+\tilde{d}_2}{\tilde{d}_1}} D(\mathbf{u}^{1:N}) \\
&= \mathcal{O}(r(N)^{c_d/\tilde{d}_1}) \left(\mathcal{O}(r(N))^{-1} D(\mathbf{u}^{1:N})^{\frac{\tilde{d}_1}{c_d+\tilde{d}_1+\tilde{d}_2}} \right)^{\frac{c_d+\tilde{d}_1+\tilde{d}_2}{\tilde{d}_1}} \\
&= o\left(r(N)^{c_d/\tilde{d}_1} \right).
\end{aligned}$$

Therefore $\|\mathcal{S}(\check{h}_{0:T}^{1:N}) - \widetilde{\mathsf{Q}}^N_{T,h_T}\|_{\mathrm{E}} = o(1)$, which concludes the proof.

B.5.3. Proof of Corollary 2

To prove the result we first construct a probability measure $\widetilde{\mathsf{Q}}^N_{T,h_T}$ such that the point set $\tilde{\mathbf{x}}^{1:N}_{0:T}$ generated by Algorithm 3 becomes, as N increases, arbitrary close to the point set $\check{\mathbf{x}}^{1:N}_{0:T}$ obtained using a smooth backward step described in Theorem 6. Then, we show that, if $\|\mathcal{S}(\check{\mathbf{x}}^{1:N}_{0:T})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}\rightarrow 0$, then$\|\mathcal{S}(\tilde{\mathbf{x}}^{1:N}_{0:T})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}\rightarrow 0$.

To this aims, assume that, for all $t\in 0:T$, the points $h_t^{1:N}$ are labelled so that $n<m \implies h_t^n<h_t^m$. (Note that the inequality is strict because, by Assumption 1 of Theorem 1, the points $\mathbf{x}_t^{1:N}$ are distinct.) Without loss of generality, assume that $h_t^1>0$ and let $h_t^0=0$ for all t.

To construct $\widetilde{\mathsf{Q}}^N_{T,h_T}$, let $\widehat{\mathsf{Q}}^N_{T,h}$ be such that $F_{\widehat{\mathsf{Q}}^N_{T,h}}$ is strictly increasing on $[0,h_T^N]$ with $F_{\widehat{\mathsf{Q}}^N_{T,h}}(h_T^n)=F_{\widehat{\mathbb{Q}}^N_{T,h}}(h_T^n)$ for all $n\in 1:N$ and, for $t\in 1:T$, let $K^N_{t,h}(h_t,\mathrm{d}h_{t-1})$ be such, for all $h_t\in[0,1)$, $F_{K^N_{t,h}}(h_t,\cdot)$ is strictly increasing on $[0,h^N_{t-1}]$ and

$$F_{K^N_{t,h}}(h_t,h^n_{t-1})=F_{\mathcal{M}^h_{t,\widehat{\mathbb{Q}}^N_{t-1,h}}}(h_t,h^n_{t-1}),\quad \forall n\in 1:N.$$

Let $\check{h}^{1:N}_{0:T}$ be as in Theorem 6 (with $\widetilde{\mathsf{Q}}^N_{T,h_T}$ constructed using the above choice of $\widehat{\mathsf{Q}}^N_{T,h}$ and $K^N_{t,h}(h_1,\mathrm{d}h_{t-1})$). We now show by a backward induction that, for any $t\in 0:T$, $\max_{n\in 1:N}\|\check{\mathbf{x}}_t^n-\tilde{\mathbf{x}}_t^n\|_\infty=o(1)$.

To see this, note that, by the construction of $\widehat{\mathsf{Q}}^N_{T,h}$,

$$|\check{h}^n_T-\tilde{h}^n_T|\leq \Delta^N_T,\quad \Delta^N_T:=\max_{n\in 1:N}|h_T^{n-1}-h_T^n|$$

where, by [17, Lemma 2], $\Delta^N_T\rightarrow 0$ as $N\rightarrow+\infty$. Hence, using the Hölder property of the Hilbert curve, this shows that $\max_{n\in 1:N}\|\check{\mathbf{x}}_T^n-\tilde{\mathbf{x}}_T^n\|_\infty=o(1)$.

Let $t\in 0:T-1$ and assume that $\max_{n\in 1:N}\|\check{\mathbf{x}}_{t+1}^n-\tilde{\mathbf{x}}_{t+1}^n\|_\infty=o(1)$. Let $w_t^n=h(\mathbf{x}_t^{\check{a}_t^n})$, where $\check{a}_t^n$ is the index selected at iteration t of Algorithm 3 obtained by replacing $\tilde{\mathbf{x}}_{t+1}^n$ by $\check{\mathbf{x}}_{t+1}^n$. Then, by the construction of $K^N_{t,h}$, $\max_{n\in 1:N}|w_t^n-\check{h}_T^n|=o(1)$.

We now want to show that $\max_{n\in 1:N}|w_t^n-\tilde{h}_T^n|=o(1)$. To simplify the notation, let $\tilde{m}_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})=m_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})G_{t_1}(\mathbf{x}_t,\mathbf{x}_{t+1})$. Then, using Assumption 4, simple computations show that, for $m\in 1:N$,

$$\begin{aligned}
|\widetilde{W}_t^m(\tilde{\mathbf{x}}_{t+1}^n)-\widetilde{W}_t^m(\check{\mathbf{x}}_{t+1}^n)|&\leq \frac{\big|W_t^m\tilde{m}_{t+1}(\mathbf{x}_t^m,\tilde{\mathbf{x}}_{t+1}^n)-W_t^m\tilde{m}_{t+1}(\mathbf{x}_t^m,\check{\mathbf{x}}_{t+1}^n)\big|}{\sum_{k=1}^N W_t^k\tilde{m}_{t+1}(\mathbf{x}_t^k,\tilde{\mathbf{x}}_{t+1}^n)}\\
&+W_t^m\tilde{m}_{t+1}(\mathbf{x}_t^m,\check{\mathbf{x}}_{t+1}^n)\frac{\Big|\sum_{k=1}^N W_t^k\tilde{m}_{t+1}(\mathbf{x}_t^k,\check{\mathbf{x}}_{t+1}^n)-\sum_{k=1}^N W_t^k\tilde{m}_{t+1}(\mathbf{x}_t^k,\tilde{\mathbf{x}}_{t+1}^n)\Big|}{\big(\sum_{k=1}^N W_t^k\tilde{m}_{t+1}(\mathbf{x}_t^k,\tilde{\mathbf{x}}_{t+1}^n)\big)\big(\sum_{k=1}^N W_t^k\tilde{m}_{t+1}(\mathbf{x}_t^k,\check{\mathbf{x}}_{t+1}^n)\big)}\\
&\leq\|G_t\|_\infty\frac{|\tilde{m}_{t+1}(\mathbf{x}_t^m,\tilde{\mathbf{x}}_{t+1}^n)-\tilde{m}_{t+1}(\mathbf{x}_t^m,\check{\mathbf{x}}_{t+1}^n)|}{N\underline{c}_t}\\
&+\|G_t\tilde{m}_{t+1}\|_\infty\frac{\sum_{k=1}^N G_t(\hat{\mathbf{x}}_{t-1}^k,\mathbf{x}_t^k)\,\big|\tilde{m}_{t+1}(\mathbf{x}_t^k,\check{\mathbf{x}}_{t+1}^n)-\tilde{m}_{t+1}(\mathbf{x}_t^k,\tilde{\mathbf{x}}_{t+1}^n)\big|}{(N\underline{c}_t)^2}.
\end{aligned}$$

Let

$$\omega_{t+1}(\delta)=\sup_{\substack{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{X}^2,\,(\mathbf{x}'_1,\mathbf{x}'_2)\in\mathcal{X}^2\\ \|\mathbf{x}_i-\mathbf{x}'_i\|_\infty\leq\delta,\,i=1,2}}|\tilde{m}_{t+1}(\mathbf{x}_1,\mathbf{x}_2)-\tilde{m}_{t+1}(\mathbf{x}'_1,\mathbf{x}'_2)|,\quad \delta>0$$

be the modulus of continuity of $\tilde{m}_{t+1}$. Then,

$$\begin{aligned}|\widetilde{W}_t^i(\tilde{\mathbf{x}}_{t+1}^n)-\widetilde{W}_t^i(\check{\mathbf{x}}_{t+1}^n)|&\leq\max_{n\in 1:N}\frac{w_{t+1}(|\tilde{\mathbf{x}}_{t+1}^n-\check{\mathbf{x}}_{t+1}^n|_\infty)}{N}\frac{\|G_t\|_\infty(\underline{c}_t+\|G_t\tilde{m}_{t+1}\|_\infty)}{\underline{c}_t^2}\\&=:\tilde{\xi}_t^N\end{aligned}$$

where, using the fact that $\tilde{m}_{t+1}$ is uniformly continuous on $\mathcal{X}^2$ (Assumption 5) and the inductive hypothesis, $\tilde{\xi}_t^N=o(N^{-1})$. Also, we know that

$$\min_{m\in 1:N}\inf_{\mathbf{x}_{t+1}\in\mathcal{X}}\widetilde{W}_t^m(\mathbf{x}_{t+1})\geq\xi_t^N:=\frac{\underline{c}_t}{N\|G_t\|\tilde{m}_{t+1}\|_\infty}.$$

Then, let N_t be such that $\tilde{\xi}_t^{N_t}<\xi_t^{N_t}$ so that, for $N\geq N_t$, we either have $\tilde{h}_t^n=w_t^n$, or $\tilde{h}_t^n=w_t^{n+1}$ or $\tilde{h}_t^n=w_t^{n-1}$. Hence, $\max_{n\in 1:N}|w_t^n-\tilde{h}_t^n|=o(1)$ and therefore $\max_{n\in 1:N}|\tilde{h}_t^n-\check{h}_t^n|=o(1)$. Finally, by, the Hölder property of the Hilbert curve, this shows that $\max_{n\in 1:N}\|\check{\mathbf{x}}_t^n-\tilde{\mathbf{x}}_t^n\|_\infty=o(1)$.

The rest of the proof follows the lines of [28, Lemma 2.5, p.15]. First, note that the above computations shows that, for any $\epsilon>0$, there exists a N_ϵ such that $\|\tilde{\mathbf{x}}_{0:T}^{1:N}-\check{\mathbf{x}}_{0:T}^{1:N}\|_\infty\leq\epsilon$ for $N\geq N_\epsilon$. Let $B=[\boldsymbol{a},\boldsymbol{b}]$, $B^+=[\boldsymbol{a},\boldsymbol{b}+\epsilon]\cap[0,1)^{T+1}$ and $B^-=[\boldsymbol{a},\boldsymbol{b}-\epsilon]$. If $\epsilon>b_i$ for at least one $i\in 1:(T+1)$, $B^-=\varnothing$. Then for $N\geq N_\epsilon$, we have

$$\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})(B^-)\leq\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(B)\leq\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})(B^+).\tag{25}$$

By the definition of the extreme metric, we have

$$\begin{aligned}\left|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})(B^+)-\widetilde{\mathbb{Q}}_T(B^+)\right|&\leq\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}},\\\left|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})(B^-)-\widetilde{\mathbb{Q}}_T(B^-)\right|&\leq\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}.\end{aligned}\tag{26}$$

Combining (25) and (26) yields:

$$\begin{cases}-\Big(\widetilde{\mathbb{Q}}_T(B)-\widetilde{\mathbb{Q}}_T(B^-)\Big)-\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}\leq\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(B)-\widetilde{\mathbb{Q}}_T(B)\\\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(B)-\widetilde{\mathbb{Q}}_T(B)\leq\Big(\widetilde{\mathbb{Q}}_T(B^+)-\widetilde{\mathbb{Q}}_T(B)\Big)+\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}.\end{cases}\tag{27}$$

Using the fact that $\widetilde{\mathbb{Q}}_T$ admits a bounded density, we have for a constant $c>0$

$$\begin{aligned}\widetilde{\mathbb{Q}}_T(B)-\widetilde{\mathbb{Q}}_T(B^-)&\leq c\lambda_{T+1}(B\setminus B^-)\leq c\,\epsilon^{T+1}\\\widetilde{\mathbb{Q}}_T(B^+)-\widetilde{\mathbb{Q}}_T(B)&\leq c\lambda_{T+1}(B^+\setminus B)\leq c\,\epsilon^{T+1}.\end{aligned}\tag{28}$$

Therefore, combining (27) and (28), we obtain, for $N\geq N_\epsilon$ and for all $B\in\mathcal{B}_{[0,1)^{T+1}}$,

$$-c\,\epsilon^{T+1}-\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}\leq\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})(B)-\widetilde{\mathbb{Q}}_T(B)\leq\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}+c\,\epsilon^{T+1}$$

and thus

$$\|\mathcal{S}(\tilde{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}\leq\|\mathcal{S}(\check{\mathbf{x}}_{0:T}^{1:N})-\widetilde{\mathbb{Q}}_T\|_{\mathrm{E}}+c\,\epsilon^{T+1}$$

and the result follows from Theorem 6.