Statistica Sinica **34** (2024), 1115-1144
doi:https://doi.org/10.5705/ss.202022.0215

PARTICLE-BASED, RAPID INCREMENTAL SMOOTHER MEETS PARTICLE GIBBS

Gabriel Cardoso*, Eric Moulines and Jimmy Olsson

Ecole polytechnique, Electrophysiology and Heart Modeling Institute and KTH Royal Institute of Technology

Abstract: The particle-based rapid incremental smoother (`PARIS`) is a sequential Monte Carlo technique that allows for efficient online approximations of expectations of additive functionals under Feynman–Kac path distributions. Under weak assumptions, the algorithm has linear computational complexity and limited memory requirements. It also comes with a number of nonasymptotic bounds and convergence results. However, being based on self-normalized importance sampling, the `PARIS` estimator is biased. This bias is inversely proportional to the number of particles, but has been found to grow linearly with the time horizon, under appropriate mixing conditions. In this work, we propose the Parisian particle Gibbs (`PPG`) sampler, which has essentially the same complexity as that of the `PARIS`, but significantly reduces the bias for a given computational complexity at the cost of a modest increase in the variance. This method is a wrapper, in the sense that it uses the `PARIS` algorithm in the inner loop of the particle Gibbs algorithm to form a bias-reduced version of the targeted quantities. We substantiate the `PPG` algorithm with theoretical results, including new bounds on the bias and variance, as well as deviation inequalities. We illustrate our theoretical results using numerical experiments that support our claims.

Key words and phrases: Bias reduction, particle filters, particle Gibbs, sequential Monte Carlo, smoothing of additive functionals, state space smoothing.

1. Introduction

Feynman–Kac formulae play a key role in many models used in statistics, physics, and many other fields; see Del Moral (2004), Del Moral (2013) and Chopin and Papaspiliopoulos (2020), and the references therein. Let $\{(\mathsf{X}_n, \mathcal{X}_n)\}_{n\in\mathbb{N}}$ be a sequence of measurable spaces and define, for every $n \in \mathbb{N}$, $\mathsf{X}_{0:n} \coloneqq \prod_{m=0}^{n} \mathsf{X}_m$ and $\mathcal{X}_{0:n} \coloneqq \bigotimes_{m=0}^{n} \mathcal{X}_m$. For a sequence $\{M_n\}_{n\in\mathbb{N}}$ of Markov kernels $M_n : \mathsf{X}_n \times \mathcal{X}_{n+1} \to [0,1]$, an initial distribution $\eta_0 \in \mathsf{M}_1(\mathcal{X}_0)$, and a sequence $\{g_n\}_{n\in\mathbb{N}}$ of bounded measurable potential functions $g_n : \mathsf{X}_n \to \mathbb{R}_+$, a sequence $\{\eta_{0:n}\}_{n\in\mathbb{N}}$ of *Feynman–Kac path measures* is defined by

$$\eta_{0:n} : \mathcal{X}_{0:n} \ni A \mapsto \frac{\gamma_{0:n}(A)}{\gamma_{0:n}(\mathsf{X}_{0:n})}, \quad n \in \mathbb{N}, \tag{1.1}$$

*Corresponding author.

where

$$\gamma_{0:n} : \mathcal{X}_{0:n} \ni A \mapsto \int \mathbb{1}_A(x_{0:n})\, \eta_0(\mathrm{d}x_0) \prod_{m=0}^{n-1} Q_m(x_m, \mathrm{d}x_{m+1}), \tag{1.2}$$

with

$$Q_m : \mathsf{X}_m \times \mathcal{X}_{m+1} \ni (x, A) \mapsto g_m(x) M_m(x, A) \tag{1.3}$$

being unnormalized kernels. By convention, $\eta_{0:0} \coloneqq \eta_0$. Note that each $\eta_{0:n}$ is a probability measure, whereas $\gamma_{0:n}$ is not normalized. For every $n \in \mathbb{N}^*$, we also define the marginal distribution $\eta_n : \mathcal{X}_n \ni A \mapsto \eta_{0:n}(\mathsf{X}_{0:n-1} \times A)$. In the context of nonlinear filtering in *general state-space hidden Markov models*(HMMs), $\eta_{0:n}$ and η_n are, the *joint smoothing* and *filter distribution*, respectively, at time n; see Del Moral (2004), Cappé, Moulines and Rydén (2005) and Chopin and Papaspiliopoulos (2020).

For most problems of practical interest, the Feynman–Kac path or marginal measures are intractable, and so is any expectation associated with the same. As a result, considerable research has been devoted to developing Monte Carlo, or *particle*, approximations of such measures. A *particle filter* approximates the marginal distribution flow $\{\eta_n\}_{n\in\mathbb{N}}$ by a sequence of occupation measures, associated with a swarm of *particles* $\{\xi_n^i\}_{i=1}^N$, $n \in \mathbb{N}$, where each particle ξ_n^i is a random draw in X_n. Particle filters revolve around two operations: a *selection step*, which duplicates or sorts out particles with large or small importance weights, respectively, and a *mutation step*, which randomly evolves the selected particles in the state space. An alternating and iterative application of selection and mutation results in a swarm of N particles that are both serially and spatially dependent. Feynman–Kac path models can also be interpreted as laws associated with a certain type of Markovian backward dynamics; this interpretation is useful, for example, for the smoothing problem in nonlinear filtering (Douc et al. (2011); Del Moral, Doucet and Singh (2010)). Several convergence results have been established for particle filters, as the number N of particles tends to infinity; see for example, Del Moral (2004), Douc and Moulines (2008), Del Moral (2013) and Chopin and Papaspiliopoulos (2020). In addition, a number of nonasymptotic results have been obtained for these methods, including bounds on their bias and L_p error, as well as exponential concentration inequalities and propagation of chaos estimates. Extensions to the backward interpretation can also be found in Douc et al. (2011) and Del Moral, Doucet and Singh (2010).

In this work, we focus on the problem of recursively computing smoothed expectations

$$\eta_{0:n} h_n = \int h_n(x_{0:n})\, \eta_{0:n}(\mathrm{d}x_{0:n}), \quad n \in \mathbb{N},$$

where we introduce the vector notation $x_{0:n} = (x_0, \ldots, x_n) \in \mathsf{X}_{0:n} \coloneqq \mathsf{X}_0 \times \cdots \times \mathsf{X}_n$ for *additive functionals* h_n of the form

$$h_n(x_{0:n}) \coloneqq \sum_{m=0}^{n-1} \tilde{h}_m(x_{m:m+1}), \quad x_{0:n} \in \mathsf{X}_{0:n}. \tag{1.4}$$

In nonlinear filtering problems, such expectations appear in the context of maximum-likelihood parameter estimation, for instance, when computing the *score function* (the gradient of the log-likelihood function) or the *expectation–maximization* (EM) surrogate; see Cappé (2001), Andrieu and Doucet (2003), Poyiadjis, Doucet and Singh (2005), Cappé (2011) and Poyiadjis, Doucet and Singh (2011). In Olsson and Westerborn (2017), the authors propose an efficient *particle-based rapid incremental smoother* (`PARIS`), with linear computational complexity in the number of particles under weak assumptions and limited memory requirements, that samples on-the-fly from the backward dynamics induced by the particle filter. An interesting feature is that it requires two or more backward draws per particle to cope with the degeneracy of the sampled trajectories and remain numerically stable in the long run, with an asymptotic variance that grows only linearly with time.

In this paper, we propose a method to reduce the bias of the `PARIS` estimator of $\eta_{0:n}h_n$. The idea is to mix the `PARIS` with a version of the *particle Gibbs* algorithm with backward sampling (Andrieu, Doucet and Holenstein (2010); Lindsten, Jordan and Schön (2014); Chopin and Singh (2015); Del Moral, Kohn and Patras (2016); Del Moral and Jasra (2018)) by introducing a conditional `PARIS` algorithm. This leads to the *Parisian particle Gibbs* (`PPG`) *algorithm*, from which we derive an upper bound on the bias that decreases inversely proportionally to the number of particles and exponentially fast with the iteration index (under assumptions guaranteeing that the particle Gibbs sampler is uniformly ergodic).

The remainder of the paper is structured as follows. In Section 2 we discuss the Feynman–Kac model, along with its backward interpretation, and introduce the particle Gibbs sampler. Our presentation is inspired by Del Moral, Kohn and Patras (2016), but differs in that it avoids the use of quotient spaces of Del Moral, Kohn and Patras (2016) and the extension of the distribution to the particle ancestral indices of Andrieu, Doucet and Holenstein (2010). In Section 3, we introduce the `PARIS` algorithm and its conditional version, and show how it can be coupled with the particle Gibbs method with backward sampling, yielding the `PPG` algorithm. In Section 4, we present the central result of this study, namely, an upper bound on the bias of the `PPG` estimator as a function of the number of particles and the iteration index of the Gibbs algorithm. In addition, we provide an upper bound on the mean-squared error (MSE). In Section 5, we provide numerical experiment to illustrate our results. In Section 6, we present the most important and original proofs. Finally, the supplementary material contain pseudocode and additional technical proofs, respectively.

Notation. Let $\mathbb{R}_+ \coloneqq [0,\infty)$, $\mathbb{R}_+^* \coloneqq (0,\infty)$, $\mathbb{N} \coloneqq \{0,1,2,\ldots\}$, and $\mathbb{N}^* \coloneqq \{1,2,3,\ldots\}$ denote the sets of nonnegative and positive real numbers and the same for integers, respectively. We denote by I_N the $N \times N$ identity matrix. For any quantities $\{a_\ell\}_{\ell=m}^n$, we denote vectors as $a_{m:n} \coloneqq (a_m,\ldots,a_n)$, and for any $(m,n) \in \mathbb{N}^2$ such that $m \le n$, we let $[\![m,n]\!] \coloneqq \{m,m+1,\ldots,n\}$. For a given measurable space $(\mathsf{X},\mathcal{X})$, where $\mathcal{X}$ is a countably generated σ-field, we denote by $\mathsf{F}(\mathcal{X})$ the set of bounded $\mathcal{X}/\mathcal{B}(\mathbb{R})$-measurable functions on X. For any $h \in \mathsf{F}(\mathcal{X})$, we let $\|h\|_\infty \coloneqq \sup_{x\in\mathsf{X}} |h(x)|$ and $\mathrm{osc}(h) \coloneqq \sup_{(x,x')\in\mathsf{X}^2} |h(x)-h(x')|$ denote the supremum and oscillator norms, respectively, of h. Let $\mathsf{M}(\mathcal{X})$ be the set of σ-finite measures on $(\mathsf{X},\mathcal{X})$, and $\mathsf{M}_1(\mathcal{X}) \subset \mathsf{M}(\mathcal{X})$ be the probability measures.

Let $(\mathsf{Y},\mathcal{Y})$ be another measurable space. A possibly unnormalized transition kernel K on $\mathsf{X}\times\mathcal{Y}$ induces two integral operators, one acting on measurable functions, and the other on measures; specifically, for $h \in \mathsf{F}(\mathcal{X}\otimes\mathcal{Y})$ and $\nu \in \mathsf{M}_1(\mathcal{X})$, define the measurable function

$$Kh : \mathsf{X} \ni x \mapsto \int h(x,y)\, K(x,\mathrm{d}y)$$

and the measure

$$\nu K : \mathcal{Y} \ni A \mapsto \int K(x,A)\, \nu(\mathrm{d}x),$$

whenever these quantities are well defined. Now, let $(\mathsf{Z},\mathcal{Z})$ be a third measurable space and L be another possibly unnormalized transition kernel on $\mathsf{Y}\times\mathcal{Z}$; we then define, with K as above, two different products of K and L, namely,

$$KL : \mathsf{X}\times\mathcal{Z} \ni (x,A) \mapsto \int L(y,A)\, K(x,\mathrm{d}y)$$

and

$$K \otimes L : \mathsf{X}\times(\mathcal{Y}\otimes\mathcal{Z}) \ni (x,A) \mapsto \iint \mathbb{1}_A(y,z)\, K(x,\mathrm{d}y)\, L(y,\mathrm{d}z),$$

whenever these are well defined. This also defines the $\otimes$ product of a kernel K on $\mathsf{X}\times\mathcal{Y}$ and a measure ν on $\mathcal{X}$, as well as of a kernel L on $\mathsf{Y}\times\mathcal{X}$ and a measure μ on $\mathcal{Y}$, as the measures

$$\nu \otimes K : \mathcal{X}\otimes\mathcal{Y} \ni A \mapsto \iint \mathbb{1}_A(x,y)\, K(x,\mathrm{d}y)\, \nu(\mathrm{d}x),$$
$$L \otimes \mu : \mathcal{X}\otimes\mathcal{Y} \ni A \mapsto \iint \mathbb{1}_A(x,y)\, L(y,\mathrm{d}x)\, \mu(\mathrm{d}y).$$

2. Particle Models

In the next sections, we discuss *many-body Feynman–Kac models*, *backward interpretations*, *conditional dual processes*, and the `PARIS` algorithm. Our presentation follows that of Del Moral, Kohn and Patras (2016) closely, but with a

different definition of the many-body extensions. We restate (in Theorem 1) a duality formula of Del Moral, Kohn and Patras (2016) relating these concepts. This formula provides a foundation for the particle Gibbs sampler described in Section 2.3 and subsequent developments.

2.1. Many-body Feynman–Kac models

In the following, we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution flow $\{\eta_m\}_{m\in\mathbb{N}}$ is intractable, in general, but can be approximated by using random samples $\boldsymbol{\xi}_m = (\xi_m^1, \dots, \xi_n^N)$, for $m \in \mathbb{N}$, of particles, where $N \in \mathbb{N}^*$ is a fixed Monte Carlo sample size and each particle ξ_m^i is an X_m-valued random variable. Such a particle approximation is based on the recursion $\eta_{m+1} = \Phi_m(\eta_m)$, for $m \in \mathbb{N}$, where Φ_m denotes the mapping

$$\Phi_m : \mathsf{M}_1(\mathcal{X}_m) \ni \eta \mapsto \frac{\eta Q_m}{\eta g_m}, \tag{2.1}$$

taking on values in $\mathsf{M}_1(\mathcal{X}_{m+1})$. In order to describe recursively the evolution of the particle population, let $m \in \mathbb{N}$ and assume that the particles $\boldsymbol{\xi}_m$ form a consistent approximation of η_m, in the sense that $\mu(\boldsymbol{\xi}_m)h$, where $\mu(\boldsymbol{\xi}_m) \coloneqq N^{-1}\sum_{i=1}^N \delta_{\xi_m^i}$ (with δ_x denoting the Dirac measure located at x) is the occupation measure formed by $\boldsymbol{\xi}_m$, serves as a proxy for $\eta_m h$ for any η_m-integrable test function h. (Under general conditions, $\mu(\boldsymbol{\xi}_m)h$ converges in probability to η_m as $N \to \infty$; see Del Moral (2004) and Chopin and Papaspiliopoulos (2020), and the references therein.) Then, in order to generate an updated particle sample approximating η_{m+1}, new particles $\boldsymbol{\xi}_{m+1} = (\xi_{m+1}^1, \dots, \xi_{m+1}^N)$ are drawn conditionally independently given $\boldsymbol{\xi}_m$ according to

$$\xi_{m+1}^i \sim \Phi_m(\mu(\boldsymbol{\xi}_m)) = \sum_{\ell=1}^N \frac{g_m(\xi_m^\ell)}{\sum_{\ell'=1}^N g_m(\xi_m^{\ell'})} M_m(\xi_m^\ell, \cdot), \quad i \in [\![1, N]\!].$$

Because this process of particle updating involves sampling from the mixture distribution $\Phi_m(\mu(\boldsymbol{\xi}_m))$, it can be decomposed into two substeps: *selection* and *mutation*. The selection step randomly chooses the ℓth mixture stratum with probability $g_m(\xi_m^\ell)/\sum_{\ell'=1}^N g_m(\xi_m^{\ell'})$, and the mutation draws a new particle ξ_{m+1}^i from the selected stratum $M_m(\xi_m^\ell, \cdot)$. In Del Moral, Kohn and Patras (2016), the term *many-body Feynman–Kac models* is related to the law of process $\{\boldsymbol{\xi}_m\}_{m\in\mathbb{N}}$. For all $m \in \mathbb{N}$, let $\mathbf{X}_m \coloneqq \mathsf{X}_m^N$ and $\boldsymbol{\mathcal{X}}_m \coloneqq \mathcal{X}_m^{\otimes N}$; then, $\{\boldsymbol{\xi}_m\}_{m\in\mathbb{N}}$ is an inhomogeneous Markov chain on $\{\mathbf{X}_m\}_{m\in\mathbb{N}}$, with transition kernels

$$\boldsymbol{M}_m : \mathbf{X}_m \times \boldsymbol{\mathcal{X}}_{m+1} \ni (\boldsymbol{x}_m, A) \mapsto \Phi_m\{\mu(\boldsymbol{x}_m)\}^{\otimes N}(A)$$

and initial distribution $\boldsymbol{\eta}_0 = \eta_0^{\otimes N}$. Now, denote $\mathbf{X}_{0:n} \coloneqq \prod_{m=0}^n \mathbf{X}_m$ and $\boldsymbol{\mathcal{X}}_{0:n} \coloneqq \bigotimes_{m=0}^n \boldsymbol{\mathcal{X}}_m$. (Here, and in the following, we use a bold symbol to stress that a

quantity is related to the many-body process.) The *many-body Feynman–Kac path model* refers to the flows $\{\boldsymbol{\gamma}_m\}_{m\in\mathbb{N}}$ and $\{\boldsymbol{\eta}_m\}_{m\in\mathbb{N}}$ of the unnormalized and normalized probability distributions, respectively, on $\{\boldsymbol{\mathcal{X}}_{0:m}\}_{m\in\mathbb{N}}$ generated by (1.1) and (1.2) for the Markov kernels $\{\boldsymbol{M}_m\}_{m\in\mathbb{N}}$, the initial distribution $\boldsymbol{\eta}_0$, the potential functions

$$\boldsymbol{g}_m : \mathbf{X}_m \ni \boldsymbol{x}_m \mapsto \mu(\boldsymbol{x}_m) g_m = \frac{1}{N}\sum_{i=1}^{N} g_m(x_m^i), \quad m \in \mathbb{N},$$

and the corresponding unnormalized transition kernels

$$\boldsymbol{Q}_m : \mathbf{X}_m \times \boldsymbol{\mathcal{X}}_{m+1} \ni (\boldsymbol{x}_m, A) \mapsto \boldsymbol{g}_m(\boldsymbol{x}_m)\boldsymbol{M}_m(\boldsymbol{x}_m, A), \quad m \in \mathbb{N}.$$

Finally, note that in the previous construction, the Markov property of the many-body Feynman–Kac model relies on the fact that each potential g_m is a function of a single state x_m only, as is the case in the standard Feynman–Kac model framework (Del Moral (2004)), and that the evolution of the particles follows the model dynamics given in (2.1) (so-called *bootstrap particle filtering*). In order to extend this to more general models (such as models where the potentials are allowed to depend on two consecutive states (Lee, Singh and Vihola (2020)) or, even more generally, where no structure at all is assumed for the unnormalized kernels (1.3) (Gloaguen, Le Corff and Olsson (2022))) and particle dynamics (such as the *auxiliary particle filtering* framework introduced in Pitt and Shephard (1999)), we need to form a Markovian many-body process with tractable dynamics by furnishing each particle with an importance weight and an index that records the particle's ancestor in the previous generation. However, to avoid this technicality and to allow for a more clear-cut presentation of the methods and theoretical analysis in the coming sections, we stay within the framework of the standard Feynman–Kac models and bootstrap-type particle filters, even though extensions to more general settings may be possible.

2.2. Backward interpretation of Feynman–Kac path flows

Suppose that each kernel Q_n, for $n \in \mathbb{N}$, defined in (1.3), has a transition density q_n with respect to some dominating measure $\lambda_{n+1} \in \mathsf{M}(\mathcal{X}_{n+1})$. Then, for $n \in \mathbb{N}$ and $\eta \in \mathsf{M}_1(\mathcal{X}_n)$, we define the *backward kernel*

$$\overleftarrow{Q}_{n,\eta} : \mathsf{X}_{n+1} \times \mathcal{X}_n \ni (x_{n+1}, A) \mapsto \frac{\int \mathbb{1}_A(x_n) q_n(x_n, x_{n+1})\, \eta(\mathrm{d}x_n)}{\int q_n(x_n', x_{n+1})\, \eta(\mathrm{d}x_n')}. \tag{2.2}$$

Now, for $n \in \mathbb{N}^*$, denoting

$$B_n : \mathsf{X}_n \times \mathcal{X}_{0:n-1} \ni (x_n, A) \mapsto \int \cdots \int \mathbb{1}_A(x_{0:n-1}) \prod_{m=0}^{n-1} \overleftarrow{Q}_{m,\eta_m}(x_{m+1}, \mathrm{d}x_m), \tag{2.3}$$

we may state the following—now classical—*backward decomposition* of the Feynman–Kac path measures, a result that plays a pivotal role in the following.

Proposition 1. *For every $n \in \mathbb{N}^*$, it holds that $\gamma_{0:n} = \gamma_n \otimes B_n$ and $\eta_{0:n} = \eta_n \otimes B_n$.*

Although the decomposition in Proposition 1 is well known (see, e.g., Del Moral, Doucet and Singh (2010); Del Moral, Kohn and Patras (2016)), we provide a proof in Section 6.1 for completeness. Using backward decomposition, we can obtain a particle approximation of a given Feynman–Kac path measure $\eta_{0:n}$ by first sampling, in an initial forward pass, particle clouds $\{\boldsymbol{\xi}_m\}_{m=0}^n$ from $\boldsymbol{\eta}_0 \otimes \boldsymbol{M}_0 \otimes \cdots \otimes \boldsymbol{M}_{n-1}$. Then, in a subsequent backward pass, we sample N conditionally independent paths $\{\tilde{\xi}_{0:n}^i\}_{i=1}^N$ from $\mathbb{B}_n(\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_n, \cdot)$, where

$$\begin{aligned} &\mathbb{B}_n : \mathbf{X}_{0:n} \times \mathcal{X}_{0:n} \ni \\ &(\boldsymbol{x}_{0:n}, A) \mapsto \int \cdots \int \mathbb{1}_A(x_{0:n}) \left\{ \prod_{m=0}^{n-1} \overleftarrow{Q}_{m,\mu(\boldsymbol{x}_m)}(x_{m+1}, \mathrm{d}x_m) \right\} \mu(\boldsymbol{x}_n)(\mathrm{d}x_n) \end{aligned} \tag{2.4}$$

is a Markov kernel describing the time-reversed dynamics induced by the particle approximations generated in the forward pass. (Here, and in the following, we use blackboard notation to denote kernels related to many-body path spaces.) Finally, $\mu(\{\tilde{\xi}_{0:n}^i\}_{i=1}^N)h$ is returned as an estimator of $\eta_{0:n}h$ for any $\eta_{0:n}$-integrable test function h. This algorithm is referred to as the *forward-filtering backward-simulation* (`FFBSi`) *algorithm* in the literature, and was introduced in Godsill, Doucet and West (2004); see also Cappé, Godsill and Moulines (2007) and Douc et al. (2011). More precisely, given the forward particles $\{\boldsymbol{\xi}_m\}_{m=0}^n$, each path $\tilde{\xi}_{0:n}^i$ is generated by first drawing $\tilde{\xi}_n^i$ uniformly from among the particles $\boldsymbol{\xi}_n$ in the previous generation, and then drawing, recursively,

$$\tilde{\xi}_m^i \sim \overleftarrow{Q}_{m,\mu(\boldsymbol{\xi}_m)}(\tilde{\xi}_{m+1}^i, \cdot) = \sum_{j=1}^N \frac{q_m(\xi_m^j, \tilde{\xi}_{m+1}^i)}{\sum_{\ell=1}^N q_m(\xi_m^\ell, \tilde{\xi}_{m+1}^i)} \delta_{\xi_m^j}; \tag{2.5}$$

that is, given $\tilde{\xi}_{m+1}^i$, $\tilde{\xi}_m^i$ is picked at random from among $\boldsymbol{\xi}_m$ based on weights proportional to $\{q_m(\xi_m^j, \tilde{\xi}_{m+1}^i)\}_{j=1}^N$. Note that in this basic formulation of the `FFBSi` algorithm, each backward-sampling operation (2.5) requires the computation of the normalising constant $\sum_{\ell=1}^N q_m(\xi_m^\ell, \tilde{\xi}_{m+1}^i)$, which implies an overall quadratic complexity of the algorithm. Still, this heavy computational burden can be eased by using an effective accept–reject technique, as discussed in Section 2.4.

2.3. Conditional dual processes and particle Gibbs

The *dual process* associated with a given Feynman–Kac model (1.1–1.2) and a given trajectory $\{z_n\}_{n\in\mathbb{N}}$, where $z_n \in \mathsf{X}_n$ for every $n \in \mathbb{N}$, is defined as the canonical Markov chain with kernels

$$\boldsymbol{M}_n[z_{n+1}] : \mathbf{X}_n \times \boldsymbol{\mathcal{X}}_{n+1} \ni (\boldsymbol{x}_n, A)$$

$$\mapsto \frac{1}{N}\sum_{i=0}^{N-1}\left[\Phi_n\{\mu(\boldsymbol{x}_n)\}^{\otimes i}\otimes\delta_{z_{n+1}}\otimes\Phi_n\{\mu(\boldsymbol{x}_n)\}^{\otimes(N-i-1)}\right](A), \tag{2.6}$$

for $n\in\mathbb{N}$, and initial distribution

$$\boldsymbol{\eta}_0\langle z_0\rangle \coloneqq \frac{1}{N}\sum_{i=0}^{N-1}\left(\eta_0^{\otimes i}\otimes\delta_{z_0}\otimes\eta_0^{\otimes(N-i-1)}\right). \tag{2.7}$$

As is clear from (2.6–2.7), given $\{z_n\}_{n\in\mathbb{N}}$, a realization $\{\boldsymbol{\xi}_n\}_{n\in\mathbb{N}}$ of the dual process is generated as follows. At time zero, the process is initialized by inserting z_0 at a randomly selected position in the vector $\boldsymbol{\xi}_0$, while drawing independently the remaining elements in the same vector from η_0. After this, the process proceeds in a Markovian manner by, given $\boldsymbol{\xi}_n$, inserting z_{n+1} at a randomly selected position in $\boldsymbol{\xi}_{n+1}$, while drawing independently the remaining elements from $\Phi_n(\mu(\boldsymbol{\xi}_n))$.

In order to describe compactly the law of the conditional dual process, we define the Markov kernel

$$\mathbb{C}_n : \mathsf{X}_{0:n}\times\boldsymbol{\mathcal{X}}_{0:n}\ni(z_{0:n},A)\mapsto\boldsymbol{\eta}_0\langle z_0\rangle\otimes\boldsymbol{M}_0\langle z_1\rangle\otimes\cdots\otimes\boldsymbol{M}_{n-1}\langle z_n\rangle(A).$$

The following result elegantly combines the underlying model (1.1–1.2), the many-body Feynman–Kac model, the backward decomposition, and the conditional dual process.

Theorem 1 (Del Moral, Kohn and Patras (2016))**.** *For all $n\in\mathbb{N}$, it holds that*

$$\mathbb{B}_n\otimes\boldsymbol{\gamma}_{0:n}=\gamma_{0:n}\otimes\mathbb{C}_n. \tag{2.8}$$

In Del Moral, Kohn and Patras (2016), each state $\boldsymbol{\xi}_n$ of the many-body process maps an outcome ω of the sample space Ω onto an *unordered set* of N elements in X_n. However, we have chosen to let each $\boldsymbol{\xi}_n$ take values in the standard *product space* X_n^N, for two reasons. First, the construction of Del Moral, Kohn and Patras (2016) requires sophisticated measure-theoretic arguments to endow such unordered sets with suitable σ-fields and appropriate measures. Second, we see no need to ignore the index order of the particles, as long as the Markovian dynamics (2.6–2.7) of the conditional dual process are symmetrized over the particle cloud. Therefore, in Section 6.2, we include our own proof of duality (2.8) for completeness. Note that the measure (2.8) on $\mathcal{X}_{0:n}\otimes\boldsymbol{\mathcal{X}}_{0:n}$ is unnormalized, but because the kernels $\mathbb{B}_n$ and $\mathbb{C}_n$ are both Markov, normalizing the identity with $\gamma_{0:n}(\mathsf{X}_{0:n})=\boldsymbol{\gamma}_{0:n}(\mathbf{X}_{0:n})$ immediately yields

$$\mathbb{B}_n\otimes\boldsymbol{\eta}0:n=\eta_{0:n}\otimes\mathbb{C}_n. \tag{2.9}$$

Because the two sides of (2.9) provide the full conditionals, it is natural to take a data-augmentation approach, and sample the target (2.9) using a two-

stage deterministic-scan Gibbs sampler (Andrieu, Doucet and Holenstein (2010); Chopin and Singh (2015)). Specifically, assume we generate a state $(\boldsymbol{\xi}_{0:n}[\ell], \zeta_{0:n}[\ell])$ comprising a dual process with an associated path on the basis of $\ell \in \mathbb{N}$ iterations of the sampler. Then, we generate the next state $(\boldsymbol{\xi}_{0:n}[\ell+1], \zeta_{0:n}[\ell+1])$ in a Markovian fashion by first sampling $\boldsymbol{\xi}_{0:n}[\ell+1] \sim \mathbb{C}_n(\zeta_{0:n}[\ell], \cdot)$, and then sampling $\zeta_{0:n}[\ell+1] \sim \mathbb{B}_n(\boldsymbol{\xi}_{0:n}[\ell+1], \cdot)$. After arbitrary initialization (and the discard of possible burn-in), this procedure produces a Markov trajectory $\{(\boldsymbol{\xi}_{0:n}[\ell], \zeta_{0:n}[\ell])\}_{\ell \in \mathbb{N}}$, and under weak additional technical conditions, this Markov chain admits (2.9) as its unique invariant distribution. In such a case, the Markov chain is ergodic (Douc et al. (2018, Chap. 5)), and the marginal distribution of the conditioning path $\zeta_{0:n}[\ell]$ converges to the target distribution $\eta_{0:n}$. Therefore, for every $h \in \mathsf{F}(\mathcal{X}_{0:n})$, it holds that $\lim_{L\to\infty} L^{-1} \sum_{\ell=1}^{L} h(\zeta_{0:n}[\ell]) = \eta_{0:n}h$, $\mathbb{P}$-a.s.. This algorithm is given in the discussion in Whiteley (2010) of the original particle Gibbs paper (Andrieu, Doucet and Holenstein (2010)); however, the justification of Whiteley (2010), involving an extension of the law targeted by the particle Gibbs sampler to the ancestral indices of particles, differs somewhat from the one presented here.

2.4. The PARIS algorithm

In the following, we assume that we are given a sequence $\{h_n\}_{n\in\mathbb{N}}$ of additive state functionals of type (1.4). Interestingly, as noted in Cappé (2011) and Del Moral, Doucet and Singh (2010), the backward decomposition allows, when applied to additive state functionals, a forward recursion for the expectations $\{\eta_{0:n}h_n\}_{n\in\mathbb{N}}$. More specifically, using the forward decomposition $h_{n+1}(x_{0:n+1}) = h_n(x_{0:n}) + \tilde{h}_n(x_n, x_{n+1})$ and the backward kernel B_{n+1} defined in (2.3), we may write, for $x_{n+1} \in \mathsf{X}_{n+1}$,

$$\begin{aligned} &B_{n+1}h_{n+1}(x_{n+1}) \\ &= \int \overleftarrow{Q}_{n,\eta_n}(x_{n+1}, \mathrm{d}x_n) \int \Big\{ h_n(x_{0:n}) + \tilde{h}_n(x_n, x_{n+1}) \Big\} B_n(x_n, \mathrm{d}x_{0:n-1}) \\ &= \overleftarrow{Q}_{n,\eta_n}(B_n h_n + \tilde{h}_n)(x_{n+1}), \end{aligned} \tag{2.10}$$

which, by Proposition 1, implies that

$$\eta_{0:n+1}h_{n+1} = \eta_{n+1}\overleftarrow{Q}_{n,\eta_n}(B_n h_n + \tilde{h}_n). \tag{2.11}$$

The marginal flow $\{\eta_n\}_{n\in\mathbb{N}}$ can be expressed recursively using the mappings $\{\Phi_n\}_{n\in\mathbb{N}}$. Thus, (2.11) provides, in principle, a basis for an online computation of $\{\eta_{0:n}h_n\}_{n\in\mathbb{N}}$. Because the marginals are generally intractable, following Del Moral, Doucet and Singh (2010), we plug particle approximations $\mu(\boldsymbol{\xi}_{n+1})$ and $\overleftarrow{Q}_{n,\mu(\boldsymbol{\xi}_n)}$ (see (2.5)) of η_{n+1} and $\overleftarrow{Q}_{n,\mu(\eta_n)}$, respectively, into the recursion (2.11). More precisely, we proceed recursively, and assume that at time n, we have a

sample $\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$ of particles with associated statistics, where each statistic β_n^i serves as an approximation of $B_n h_n(\xi_n^i)$. Then evolving the particle cloud according to $\boldsymbol{\xi}_{n+1} \sim \boldsymbol{M}_n(\boldsymbol{\xi}_n, \cdot)$ and updating the statistics using (2.10), with $\overleftarrow{Q}_{n,\eta_n}$ replaced by $\overleftarrow{Q}_{n,\mu(\boldsymbol{\xi}_n)}$, yields the particle-wise recursion

$$\beta_{n+1}^i = \sum_{\ell=1}^N \frac{q_n(\xi_n^\ell, \xi_{n+1}^i)}{\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)} \left\{ \beta_n^\ell + \tilde{h}_n(\xi_n^\ell, \xi_{n+1}^i) \right\}, \quad i \in [\![1, N]\!], \tag{2.12}$$

and, finally, the estimator

$$\mu(\boldsymbol{\beta}_n)(\mathrm{id}) = \frac{1}{N} \sum_{i=1}^N \beta_n^i \tag{2.13}$$

of $\eta_{0:n} h_n$, where we set $\boldsymbol{\beta}_n \coloneqq (\beta_n^1, \dots, \beta_n^N)$, for $i \in [\![1, N]\!]$, and id is the identity mapping. The procedure is initialized by simply letting $\beta_0^i = 0$, for all $i \in [\![1, N]\!]$. Note that (2.13) provides a particle interpretation of the backward decomposition in Proposition 1. This algorithm is a special case of the *forward-filtering backward-smoothing* (`FFBSm`) *algorithm* (see Andrieu and Doucet (2003); Godsill, Doucet and West (2004); Douc et al. (2011); Särkkä (2013)) for additive functionals satisfying (1.4). It allows for online processing of the sequence $\{\eta_{0:n} h_n\}_{n \in \mathbb{N}}$, but also has the appealing property that only the current particles $\boldsymbol{\xi}_n$ and statistics $\boldsymbol{\beta}_n$ need to be stored in memory. However, because each update (2.12) requires a summation of N terms, the scheme has an overall *quadratic* complexity in the number of particles, leading to a computational bottleneck in applications to complex models that require large particle sample sizes N.

To avoid the computational burden of this forward-only implementation of `FFBSm`, the `PARIS` algorithm Olsson and Westerborn (2017) updates the statistics $\boldsymbol{\beta}_n$ by replacing each sum (2.12) with the Monte Carlo estimate

$$\beta_{n+1}^i = \frac{1}{M} \sum_{j=1}^M \left\{ \tilde{\beta}_n^{i,j} + \tilde{h}_n(\tilde{\xi}_n^{i,j}, \xi_{n+1}^i) \right\}, \quad i \in [\![1, N]\!], \tag{2.14}$$

where $\{(\tilde{\xi}_n^{i,j}, \tilde{\beta}_n^{i,j})\}_{j=1}^M$ are drawn randomly from among $\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$ with replacement, by assigning $(\tilde{\xi}_n^{i,j}, \tilde{\beta}_n^{i,j})$ the value of $(\xi_n^\ell, \beta_n^\ell)$ with probability $q_n(\xi_n^\ell, \xi_{n+1}^i) / \sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)$, and the Monte Carlo sample size $M \in \mathbb{N}^*$ is much smaller than N (say, less than five). Formally,

$$\{(\tilde{\xi}_n^{i,j}, \tilde{\beta}_n^{i,j})\}_{j=1}^M \sim \left\{ \sum_{\ell=1}^N \frac{q_n(\xi_n^\ell, \xi_{n+1}^i)}{\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)} \delta_{(\xi_n^\ell, \beta_n^\ell)} \right\}^{\otimes M}, \quad i \in [\![1, N]\!].$$

The resulting procedure, summarized in Algorithm 1, allows for online processing with constant memory requirements, because it only needs to store the current particle cloud and the estimated auxiliary statistics at each iteration. Moreover,

when the Markov transition densities of the model can be uniformly bounded, that is, there exists, for every $n \in \mathbb{N}$, an upper bound $\bar{\sigma}_n > 0$ such that for all $(x_n, x_{n+1}) \in \mathsf{X}_n \times \mathsf{X}_{n+1}$, $m_n(x_n, x_{n+1}) \leq \bar{\sigma}_n$ (a weak assumption satisfied for most models of interest), then we can generate a sample $(\tilde{\xi}_n^{i,j}, \beta_n^{i,j})$ by drawing, with replacement and until acceptance, candidates $(\tilde{\xi}_n^{i,*}, \tilde{\beta}_n^{i,*})$ from $\{(\xi_n^i, \beta_n^i)\}_{i=1}^N$ based on the normalized particle weights $\{g_n(\xi_n^\ell)/\sum_{\ell'} g_n(\xi_n^{\ell'})\}_{\ell=1}^N$ (obtained as a by-product in the generation of $\boldsymbol{\xi}_{n+1}$), and accepting the same with probability $m_n(\tilde{\xi}_n^{i,*}, \xi_{n+1}^i)/\bar{\sigma}_n$. Because this sampling procedure bypasses the calculation of the normalizing constant $\sum_{\ell'=1}^N q_n(\xi_n^{\ell'}, \xi_{n+1}^i)$ of the targeted categorical distribution, it yields an overall $\mathcal{O}(MN)$ complexity of the algorithm; see (Douc et al. (2011)) for details.

Increasing M improves the accuracy of the algorithm at the cost of additional computational complexity.

As shown in Olsson and Westerborn (2017), there is a qualitative difference between the cases $M = 1$ and $M \geq 2$, and the latter is required to keep the `PARIS` numerically stable. More precisely, in the latter case, it can be shown that the `PARIS` estimator $\mu(\boldsymbol{\beta}_n)$ satisfies, as N tends to infinity while M is held fixed, a central limit theorem (CLT) at the rate $\sqrt{N}$, with an n-normalized asymptotic variance of order $\mathcal{O}(1 - 1/(M-1))$. As is clear from this bound, using a large M only wastes computational work, and setting M to two or three typically works well in practice.

3. The `PPG` Sampler

We now introduce the `PPG` *algorithm*. For all $n \in \mathbb{N}^*$, let $\mathsf{Y}_n \coloneqq \mathsf{X}_{0:n} \times \mathbb{R}$ and $\mathcal{Y}_n \coloneqq \mathcal{X}_{0:n} \otimes \mathcal{B}(\mathbb{R})$. Moreover, let $\mathsf{Y}_0 \coloneqq \mathsf{X}_0 \times \{0\}$ and $\mathcal{Y}_0 \coloneqq \mathcal{X}_0 \otimes \{\{0\}, \emptyset\}$. An element of Y_n is always denoted by $y_n = (x_{0:n|n}, b_n)$. The `PPG` sampler includes, as a key ingredient, a *conditional `PARIS` step*, that recursively updates a set of Y_n-valued random variables $v_n^i \coloneqq (\xi_{0:n|n}^i, \beta_n^i)$, for $i \in [\![1, N]\!]$. Let $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ denote the corresponding many-body process, with each $\boldsymbol{v}_n \coloneqq ((\xi_{0:n|n}^1, \beta_n^1), \ldots, (\xi_{0:n|n}^N, \beta_n^N))$ taking on values in the space $\mathbf{Y}_n \coloneqq \mathsf{Y}_n^N$, which we furnish with a σ-field $\boldsymbol{\mathcal{Y}}_n \coloneqq \mathcal{Y}_n^{\otimes N}$. The space $\mathbf{Y}_0$ and the corresponding σ-field $\boldsymbol{\mathcal{Y}}_0$ are defined accordingly. For every $n \in \mathbb{N}$, we write $\boldsymbol{\xi}_{0:n|n} = (\xi_{0:n|n}^1, \ldots, \xi_{0:n|n}^N)$ for the collection of paths in $\boldsymbol{v}_n$, and $\boldsymbol{\xi}_{n|n} = (\xi_n^1, \ldots, \xi_n^N)$ for the collection of end points of the same.

In the following, we let $n \in \mathbb{N}$ be a fixed time horizon, and describe in detail how the `PPG` approximates $\eta_{0:n}h_n$ iteratively. In short, at each iteration ℓ, and given an input conditional path $\zeta_{0:n}[\ell]$, the `PPG` produces a many-body system $\boldsymbol{v}_n[\ell+1]$ by using a series of conditional `PARIS` operations. Then, an updated path $\zeta_{0:n}[\ell+1]$, which serves as input at the next iteration, is generated by picking one of the paths $\boldsymbol{\xi}_{0:n|n}[\ell+1]$ in $\boldsymbol{v}_n[\ell+1]$ at random. At each iteration, the produced statistics $\boldsymbol{\beta}_n$ (in $\boldsymbol{v}_n$) provide an approximation of $\eta_{0:n}h_n$, according to (2.13).

More precisely, given a path $\zeta_{0:n}[\ell]$, the conditional PARIS operations are executed as follows. In the initial step, $\boldsymbol{\xi}_{0|0}[\ell+1]$ are drawn from $\boldsymbol{\eta}_0\langle\zeta_0[\ell]\rangle$ defined in (2.7), and $\upsilon_0^i[\ell+1] \leftarrow (\xi_{0|0}^i[\ell+1], 0)$, for all $i \in [\![1, N]\!]$; then, recursively, for $m \in [\![0, n]\!]$, assuming access to $\boldsymbol{\upsilon}_m[\ell+1]$, we

(1) generate an updated particle cloud $\boldsymbol{\xi}_{m+1}[\ell+1] \sim \boldsymbol{M}_m\langle\zeta_{m+1}[\ell]\rangle(\boldsymbol{\xi}_{m|m}[\ell+1], \cdot)$,

(2) pick at random, for each $i \in [\![1, N]\!]$, an ancestor path with associated statistics $(\tilde{\xi}_{0:m}^{i,1}[\ell+1], \tilde{\beta}_m^{i,1}[\ell+1])$ from among $\boldsymbol{\upsilon}_m[\ell+1]$ by drawing

$$(\tilde{\xi}_{0:m}^{i,1}[\ell+1], \tilde{\beta}_m^{i,1}[\ell+1]) \sim \sum_{s=1}^{N} \frac{q_m(\xi_{m|m}^s[\ell+1], \xi_{m+1}^i[\ell+1])}{\sum_{s'=1}^{N} q_m(\xi_{m|m}^{s'}[\ell+1], \xi_{m+1}^i[\ell+1])} \delta_{\upsilon_m^s[\ell+1]},$$

(3) pick at random, for each $i \in [\![1, N]\!]$, with replacement, $M-1$ ancestor particles and associated statistics $\{(\tilde{\xi}_m^{i,j}[\ell+1], \tilde{\beta}_m^{i,j}[\ell+1])\}_{j=2}^M$ at random from $\{(\xi_{m|m}^s[\ell+1], \beta_m^s[\ell+1])\}_{s=1}^N$ according to

$$\{(\tilde{\xi}_m^{i,j}[\ell+1], \tilde{\beta}_m^{i,j}[\ell+1])\}_{j=2}^M \sim \left(\sum_{s=1}^{N} \frac{q_m(\xi_{m|m}^s[\ell+1], \xi_{m+1}^i[\ell+1])}{\sum_{s'=1}^{N} q_m(\xi_{m|m}^{s'}[\ell+1], \xi_{m+1}^i[\ell+1])} \delta_{(\xi_{m|m}^s[\ell+1], \beta_m^s[\ell+1])} \right)^{\otimes(M-1)},$$

(4) set, for all $i \in [\![1, N]\!]$, $\xi_{0:m+1|m+1}^i[\ell+1] \leftarrow (\tilde{\xi}_{0:m}^{i,1}[\ell+1], \xi_{m+1}^i[\ell+1])$ and $\upsilon_{m+1}^i[\ell+1] \leftarrow (\xi_{0:m+1|m+1}^i[\ell+1], \beta_{m+1}^i[\ell+1])$, where

$$\beta_{m+1}^i[\ell+1] \leftarrow M^{-1} \sum_{j=1}^{M} \left(\tilde{\beta}_m^{i,j}[\ell+1] + \tilde{h}_m(\tilde{\xi}_m^{i,j}[\ell+1], \xi_{m+1}^i[\ell+1]) \right).$$

This conditional PARIS procedure is summarized in pseudocode in Algorithm 2 in Section B.

In addition to recursively propagating the statistics $\{\boldsymbol{\beta}_m[\ell+1]\}_{m=0}^n$ to form the final estimator, this scheme also recursively propagates the trajectories $\{\boldsymbol{\xi}_{0:m|m}[\ell+1]\}_{m=0}^n$ used as a pool of candidates for the updated conditional path $\zeta_{0:n}[\ell+1]$. Once we have the set $\boldsymbol{\upsilon}_n[\ell+1]$ of trajectories and associated statistics formed using n recursive conditional PARIS updates, we draw an updated path $\zeta_{0:n}[\ell+1]$ from $\mu(\boldsymbol{\xi}_{0:n|n}[\ell+1])$ (*i.e.*, uniformly among the elements of $\boldsymbol{\xi}_{0:n|n}[\ell+1]$). As a result, the updated conditional path $\zeta_{0:n}[\ell+1]$ and the statistics $\boldsymbol{\beta}_n[\ell+1]$ are statistically intertwined conditionally on the conditional dual particle process underpinning the algorithm. The main reason for this is to avoid computational waste. By letting the updated conditional path $\zeta_{0:n}[\ell+1]$ be formed by reusing the backward samples from those generated to form the statistics $\boldsymbol{\beta}_n[\ell+1]$ included in the estimator, our procedure optimizes available computational resources. The full PPG is summarized in pseudocode in Algorithm 3 in Section B.

The following Markov kernels play an instrumental role in the following. For a given path $\{z_m\}_{m\in\mathbb{N}}$, the conditional PARIS update in Algorithm 2 defines an inhomogeneous Markov chain on the spaces $\{(\mathbf{Y}_m, \boldsymbol{\mathcal{Y}}_m)\}_{m\in\mathbb{N}}$ with kernels

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

where

$$\begin{aligned}
&\boldsymbol{S}_m : \mathbf{Y}_m \times \mathbf{X}_{m+1} \times \boldsymbol{\mathcal{Y}}_{m+1} \ni (\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, A) \qquad (3.1)\\
&\mapsto \int \cdots \int \mathbb{1}_A \left(\left\{ \left((\tilde{x}^{i,1}_{0:m}, x^i_{m+1}), \frac{1}{M} \sum_{j=1}^{M} \left(\tilde{b}^{i,j}_m + \tilde{h}_m(\tilde{x}^{i,j}_m, x^i_{m+1}) \right) \right) \right\}_{i=1}^{N} \right)\\
&\quad \times \prod_{i=1}^{N} \left(\sum_{\ell=1}^{N} \frac{q_m(x^\ell_{m|m}, x^i_{m+1})}{\sum_{\ell'=1}^{N} q_m(x^{\ell'}_{m|m}, x^i_{m+1})} \delta_{y^\ell_m}\, \mathrm{d}(\tilde{x}^{i,1}_{0:m}, \tilde{b}^{i,1}_m) \right.\\
&\quad \left. \times \left\{ \sum_{\ell=1}^{N} \frac{q_m(x^\ell_{m|m}, x^i_{m+1})}{\sum_{\ell'=1}^{N} q_m(x^{\ell'}_{m|m}, x^i_{m+1})} \delta_{(x^\ell_{m|m}, b^\ell_m)} \right\}^{\otimes(M-1)} \mathrm{d}(\tilde{x}^{i,2}_m, \tilde{b}^{i,2}_m, \ldots, \tilde{x}^{i,M}_m, \tilde{b}^{i,M}_m) \right).
\end{aligned}$$

In addition, we introduce the joint law

$$\begin{aligned}
&\mathbb{S}_n : \mathbf{X}_{0:n} \times \boldsymbol{\mathcal{Y}}_n \ni (\boldsymbol{x}_{0:n}, A)\\
&\mapsto \int \cdots \int \mathbb{1}_A(\boldsymbol{y}_n)\, \boldsymbol{S}_0(\mathbf{J}\boldsymbol{x}_0, \boldsymbol{x}_1, \mathrm{d}\boldsymbol{y}_1) \prod_{m=1}^{n-1} \boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, \mathrm{d}\boldsymbol{y}_{m+1}), \qquad (3.2)
\end{aligned}$$

where we define $\mathbf{J} \coloneqq \mathrm{I}_N \otimes (0,1)^\intercal$.

The kernel $\mathbb{S}_n$ can be viewed as a *superincumbent sampling kernel* that describes the distribution of the output $\boldsymbol{\upsilon}_n$ generated by a sequence of PARIS iterations when the many-body process $\{\boldsymbol{\xi}_m\}_{m=0}^n$ associated with the underlying particle filter is given. This allows us to describe the PPG alternatively as follows: given $\zeta_{0:n}[\ell]$, draw $\boldsymbol{\xi}_{0:n}[\ell+1] \sim \mathbb{C}_n(\zeta_{0:n}[\ell], \cdot)$; then, draw $\boldsymbol{\upsilon}_n[\ell+1] \sim \mathbb{S}_n(\boldsymbol{\xi}_{0:n}[\ell+1], \cdot)$ and pick a trajectory $\zeta_{0:n}[\ell+1]$ from $\boldsymbol{\xi}_{0:n|n}[\ell+1]$ at random. The following proposition, establishes that the conditional distribution of $\zeta_{0:n}[\ell+1]$ given $\boldsymbol{\xi}_{0:n}[\ell+1]$ coincides, as expected, with the particle-induced backward dynamics $\mathbb{B}_n$.

Proposition 2. *For all $n \in \mathbb{N}^*$, $N \in \mathbb{N}^*$, $\boldsymbol{x}_{0:n} \in \mathbf{X}_{0:n}$, and $h \in \mathsf{F}(\mathcal{X}_{0:n})$,*

$$\int \mathbb{S}_n(\boldsymbol{x}_{0:n}, \mathrm{d}\boldsymbol{y}_n)\, \mu(\boldsymbol{x}_{0:n|n}) h = \mathbb{B}_n h(\boldsymbol{x}_{0:n}).$$

Finally, we define the Markov kernel induced by the PPG, as well as the extended probability distribution targeted by the same. For this purpose, we introduce the extended measurable space $(\mathbf{E}_n, \boldsymbol{\mathcal{E}}_n)$, with

$$\mathbf{E}_n \coloneqq \mathbf{Y}_n \times \mathsf{X}_{0:n}, \quad \boldsymbol{\mathcal{E}}_n \coloneqq \boldsymbol{\mathcal{Y}}_n \otimes \mathcal{X}_{0:n}.$$

The PPG described in Algorithm 3 defines a Markov chain on $(\mathbf{E}_n, \boldsymbol{\mathcal{E}}_n)$ with the Markov transition kernel

$$\begin{aligned}&\mathbb{K}_n : \mathbf{E}_n \times \boldsymbol{\mathcal{E}}_n \ni (\boldsymbol{y}_n, z_{0:n}, A)\\ &\mapsto \iiint \mathbb{1}_A(\tilde{\boldsymbol{y}}_n, \tilde{z}_{0:n})\, \mathbb{C}_n(z_{0:n}, \mathrm{d}\tilde{\boldsymbol{x}}_{0:n})\, \mathbb{S}_n(\tilde{\boldsymbol{x}}_{0:n}, \mathrm{d}\tilde{\boldsymbol{y}}_n)\, \mu(\tilde{\boldsymbol{x}}_{0:n|n})(\mathrm{d}\tilde{z}_{0:n}).\end{aligned}$$

Note that the values of $\mathbb{K}_n$ defined above do not depend on $\boldsymbol{y}_n$, but only on $(z_{0:n}, A)$. For any given initial distribution $\xi \in \mathsf{M}_1(\mathcal{X}_{0:n})$, let $\mathbb{P}_\xi$ be the distribution of the canonical Markov chain induced by the kernel $\mathbb{K}_n$ and the initial distribution ξ. In the special case where $\xi = \delta_{z_{0:n}}$, for some given path $z_{0:n} \in \mathsf{X}_{0:n}$, we use the short-hand notation $\mathbb{P}_{\delta_{z_{0:n}}} = \mathbb{P}_{z_{0:n}}$. In addition, denote by

$$\begin{aligned}&K_n : \mathsf{X}_{0:n} \times \mathcal{X}_{0:n} \ni (z_{0:n}, A)\\ &\mapsto \iiint \mathbb{1}_A(\tilde{z}_{0:n})\, \mathbb{C}_n(z_{0:n}, \mathrm{d}\tilde{\boldsymbol{x}}_{0:n})\, \mathbb{S}_n(\tilde{\boldsymbol{x}}_{0:n}, \mathrm{d}\tilde{\boldsymbol{y}}_n)\, \mu(\tilde{\boldsymbol{x}}_{0:n|n})(\mathrm{d}\tilde{z}_{0:n})\end{aligned}$$

the path-marginalized version of $\mathbb{K}_n$. By Proposition 2, it holds that $K_n = \mathbb{C}_n\mathbb{B}_n$, which shows that K_n coincides with the Markov transition kernel of the backward-sampling-based particle Gibbs sampler discussed in Section 2.3.

Finally, in order to prepare for the statement of our theoretical results on the PPG, we need to introduce the following Feynman–Kac path model *with a frozen path*. More precisely, for a given path $z_{0:n} \in \mathsf{X}_{0:n}$, define, for every $m \in [\![0, n-1]\!]$, the unnormalized kernel

$$Q_m\langle z_{m+1}\rangle : \mathsf{X}_m \times \mathcal{X}_{m+1} \ni (x_m, A) \mapsto \left(1 - \frac{1}{N}\right) Q_m(x_m, A) + \frac{1}{N} g_m(x_m)\, \delta_{z_{m+1}}(A)$$

and the initial distribution $\eta_0\langle z_0\rangle : \mathcal{X}_0 \ni A \mapsto (1 - 1/N)\eta_0(A) + \delta_{z_0}(A)/N$. Given these quantities, define, for $m \in [\![0, n]\!]$, $\gamma_m\langle z_{0:m}\rangle \coloneqq \eta_0\langle z_0\rangle Q_0\langle z_1\rangle \cdots Q_{m-1}\langle z_m\rangle$, and its normalized counterpart $\eta_m\langle z_{0:m}\rangle \coloneqq \gamma_m\langle z_{0:m}\rangle / \gamma_m\langle z_{0:m}\rangle \mathbb{1}_{\mathsf{X}_{0:m}}$. Finally, we introduce, for $m \in [\![0, n]\!]$, the kernels

$$\begin{aligned}&B_m\langle z_{0:m-1}\rangle : \mathsf{X}_m \times \mathcal{X}_{0:m-1} \ni (x_m, A)\\ &\mapsto \int \cdots \int \mathbb{1}_A(x_{0:n-1}) \prod_{m=0}^{n-1} \overleftarrow{Q}_{m, \eta_m\langle z_{0:m}\rangle}(x_{m+1}, \mathrm{d}x_m)\end{aligned}$$

and the path model $\eta_{0:m}\langle z_{0:m}\rangle \coloneqq B_m\langle z_{0:m-1}\rangle \otimes \eta_m\langle z_{0:m}\rangle$.

4. Main Results

4.1. Theoretical results

In this section, we establish our main result, namely, the exponentially contracting bias bound stated in Theorem 2. This result is proved under the following strong mixing assumptions, which are standard in the literature (see Del Moral (2004); Douc and Moulines (2008); Del Moral (2013); Del Moral, Kohn and Patras (2016)):

Assumption 1 (**Strong mixing**). *For every $n \in \mathbb{N}$, there exist $\underline{\tau}_n$, $\bar{\tau}_n$, $\underline{\sigma}_n$, and $\bar{\sigma}_n$ in $\mathbb{R}_+^*$ such that*

(i) $\underline{\tau} \leq g_n(x_n) \leq \bar{\tau}_n$ *for every* $x_n \in \mathsf{X}_n$,

(ii) $\underline{\sigma}_n \leq m_n(x_n, x_{n+1}) \leq \bar{\sigma}_n$ *for every* $(x_n, x_{n+1}) \in \mathsf{X}_{n:n+1}$.

Under Assumption 1, define, for every $n \in \mathbb{N}$,

$$\rho_n \coloneqq \max_{m \in [\![0,n]\!]} \frac{\bar{\tau}_m \bar{\sigma}_m}{\underline{\tau}_m \underline{\sigma}_m} \tag{4.1}$$

and, for every $n \in \mathbb{N}$ and $N \in \mathbb{N}^*$ such that $N > N_n \coloneqq (1+5\rho_n^2 n/2) \vee 2n(1+2\rho_n^2)$,

$$\kappa_{N,n} \coloneqq 1 - \frac{1 - (1 + 5n\rho_n^2/2)/N}{1 + 4n(1 + 2\rho_n^2)/N}. \tag{4.2}$$

Note that $\kappa_{N,n} \in (0,1)$, for all N and n, as above.

Theorem 2. *Assume Assumption 1. Then, for every $n \in \mathbb{N}$, there exist c_n^{bias}, c_n^{mse}, and c_n^{cov} in $\mathbb{R}_+^*$ such that for every $M \in \mathbb{N}^*$, $\xi \in \mathsf{M}_1(\mathcal{X}_{0:n})$, $\ell \in \mathbb{N}^*$, $s \in \mathbb{N}^*$, and $N \in \mathbb{N}^*$ such that $N > N_n$,*

$$\left| \mathbb{E}_\xi \left[\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) \right] - \eta_{0:n} h_n \right| \leq \mathsf{c}_n^{bias} \left(\sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty \right) N^{-1} \kappa_{N,n}^\ell, \tag{4.3}$$

$$\mathbb{E}_\xi \left[\left(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n} h_n \right)^2 \right] \leq \mathsf{c}_n^{mse} \left(\sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty \right)^2 N^{-1}, \tag{4.4}$$

$$\begin{aligned} &\left| \mathbb{E}_\xi \left[\left(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n} h_n \right) \left(\mu(\boldsymbol{\beta}_n[\ell+s])(\mathrm{id}) \ - \eta_{0:n} h_n \right) \right] \right| \\ &\leq \mathsf{c}_n^{cov} \left(\sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty \right)^2 N^{-3/2} \kappa_{N,n}^s. \end{aligned} \tag{4.5}$$

The constants c_n^{bias}, c_n^{mse}, and c_n^{cov} are given explicitly in the proof. Because we focus on the dependence on N and the index ℓ, we make no attempt to optimize the dependence of these constants on n in our proofs; nevertheless, we believe that it is possible to prove, under the stated assumptions, that this dependence is linear. The proof of the bound in Theorem 2 is based on four key ingredients. The

first is the following unbiasedness property of the PARIS under the many-body Feynman–Kac path model.

Theorem 3. *For every $n \in \mathbb{N}$, $N \in \mathbb{N}^*$, and $\ell \in \mathbb{N}^*$,*

$$\begin{aligned}\mathbb{E}_{\eta_{0:n}}\left[\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id})\right] &= \int \eta_{0:n}\mathbb{C}_n\mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id}) \\ &= \int \boldsymbol{\eta}_{0:n}\mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id}) = \eta_{0:n}h_n.\end{aligned}$$

The proof of Theorem 3 is found in Section 6.3. The second is the uniform geometric ergodicity of the particle Gibbs with backward sampling established in Del Moral and Jasra (2018).

Theorem 4. *Assume Assumption* 1. *Then, for every $n \in \mathbb{N}$, $(\mu,\nu) \in \mathsf{M}_1(\mathcal{X}_{0:n})^2$, $\ell \in \mathbb{N}^*$, and $N \in \mathbb{N}^*$ such that $N > N_n$, $\|\mu K_n^\ell - \nu K_n^\ell\|_{\mathrm{TV}} \leq \kappa_{N,n} n N^\ell$, where $\kappa_{N,n}$ is defined in* (4.2).

As a third ingredient, we require the following uniform exponential concentration inequality of the conditional PARIS with respect to the frozen-path Feynman–Kac model defined in the previous section.

Theorem 5. *For every $n \in \mathbb{N}$, there exist $\mathsf{c}_n > 0$ and $\mathsf{d}_n > 0$ such that for every $M \in \mathbb{N}^*$, $z_{0:n} \in \mathsf{X}_{0:n}$, $N \in \mathbb{N}^*$, and $\varepsilon > 0$,*

$$\begin{aligned}&\int \mathbb{C}_n\mathbb{S}_n(z_{0:n},\mathrm{d}\boldsymbol{b}_n)\mathbb{1}\left\{|\mu(\boldsymbol{b}_n)(\mathrm{id}) - \eta_{0:n}\langle z_{0:n}\rangle h_n| \geq \varepsilon\right\} \\ &\leq \mathsf{c}_n \exp\left(-\frac{\mathsf{d}_n N\varepsilon^2}{2(\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty)^2}\right).\end{aligned}$$

The proof of Corollary 5 is found in Section C.2, and is based on arguments similar to those used in the proofs of Olsson and Westerborn (2017, Thm. 1) and Douc et al. (2011, Thm. 5) in the framework of the conditional dual process. Corollary 5 implies, in turn, the following conditional variance bound.

Proposition 3. *For every $n \in \mathbb{N}$, $M \in \mathbb{N}^*$, $z_{0:n} \in \mathsf{X}_{0:n}$, and $N \in \mathbb{N}^*$,*

$$\int \mathbb{C}_n\mathbb{S}_n(z_{0:n},\mathrm{d}\boldsymbol{b}_n)\,|\mu(\boldsymbol{b}_n)(\mathrm{id}) - \eta_{0:n}\langle z_{0:n}\rangle h_n|^2 \leq \frac{\mathsf{c}_n}{\mathsf{d}_n}\left(\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty\right)^2 N^{-1}.$$

Using Corollary 3, we deduce, in turn, the following bias bound, the proof is postponed to C.4.

Proposition 4. *For every $n \in \mathbb{N}$, there exists $\bar{\mathsf{c}}_n^{bais} > 0$ such that for every $M \in \mathbb{N}^*$, $z_{0:n} \in \mathsf{X}_{0:n}$, and $N \in \mathbb{N}^*$,*

$$\left|\int \mathbb{C}_n\mathbb{S}_n(z_{0:n},\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id}) - \eta_{0:n}\langle z_{0:n}\rangle h_n\right| \leq \bar{\mathsf{c}}_n^{bais}\left(\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty\right)N^{-1}.$$

A fourth and last ingredient in the proof of Theorem 2 is the following bound on the discrepancy between the additive expectations under the original and frozen-path Feynman–Kac models. This bound is established using novel results in Gloaguen, Le Corff and Olsson (2022). More precisely, because for every $m \in \mathbb{N}$, $(x, z) \in \mathsf{X}_m^2$, $N \in \mathbb{N}^*$, and $h \in \mathsf{F}(\mathcal{X}_{m+1})$, using Assumption 1,

$$|Q_m\langle z\rangle h(x) - Q_m h(x)| \leq \frac{1}{N}\|g_m\|_\infty \|h\|_\infty \leq \frac{1}{N}\bar{\tau}_m \|h\|_\infty,$$

applying Gloaguen, Le Corff and Olsson (2022, Thm. 4.3) yields the following.

Proposition 5. *Assume Assumption* 1. *Then, there exists* $\mathsf{c} > 0$ *such that for every* $n \in \mathbb{N}$, $N \in \mathbb{N}$, *and* $z_{0:n} \in \mathsf{X}_{0:n}$,

$$|\eta_{0:n}\langle z_{0:n}\rangle h_n - \eta_{0:n} h_n| \leq \mathsf{c} N^{-1} \sum_{m=0}^{n-1} \|\tilde{h}_m\|_\infty.$$

In addition, we assume $\sup_{n\in\mathbb{N}} \|\tilde{h}_n\|_\infty < \infty$ yields an $\mathcal{O}(n/N)$ bound in Proposition 5.

Finally, by combining these ingredients, we are now ready to present a proof of Theorem 2.

Proof of Theorem 2. Write, using the tower property,

$$\mathbb{E}_\xi\left[\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id})\right] = \mathbb{E}_\xi\left[\mathbb{E}_{\zeta_{0:n}[\ell]}\left[\mu(\boldsymbol{\beta}_n[0])(\mathrm{id})\right]\right] = \int \xi K_n^\ell \mathbb{C}_n \mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\, \mu(\boldsymbol{b}_n)(\mathrm{id}).$$

Thus, by the unbiasedness property in Theorem 3,

$$\begin{aligned}
&|\mathbb{E}_\xi\left[\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id})\right] - \eta_{0:n} h_n| \\
&= \left| \int \xi K_n^\ell \mathbb{C}_n \mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\, \mu(\boldsymbol{b}_n)(\mathrm{id}) - \int \eta_{0:n} \mathbb{C}_n \mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\, \mu(\boldsymbol{b}_n)(\mathrm{id}) \right| \\
&\leq \left\| \xi K_n^\ell - \eta_{0:n} \right\|_{\mathrm{TV}} \operatorname{osc}\left(\int \mathbb{C}_n \mathbb{S}_n(\cdot, \mathrm{d}\boldsymbol{b}_n)\, \mu(\boldsymbol{b}_n)(\mathrm{id}) \right),
\end{aligned}$$

where, by Theorem 4, $\|\xi K_n^\ell - \eta_{0:n}\|_{\mathrm{TV}} \leq \kappa_{N,n}^\ell$. Moreover, to derive an upper bound on the oscillation, we consider the decomposition

$$\begin{aligned}
&\operatorname{osc}\left(\int \mathbb{C}_n \mathbb{S}_n(\cdot, \mathrm{d}\boldsymbol{b}_n)\, \mu(\boldsymbol{b}_n)(\mathrm{id}) \right) \\
&\leq 2\left(\left\| \int \mathbb{C}_n \mathbb{S}_n(\cdot, \mathrm{d}\boldsymbol{b}_n)\, \mu(\boldsymbol{b}_n)(\mathrm{id}) - \eta_{0:n}\langle\cdot\rangle h_n \right\|_\infty + \left\| \eta_{0:n}\langle\cdot\rangle h_n - \eta_{0:n} h_n \right\|_\infty \right),
\end{aligned}$$

where the two terms on the right-hand side can be bounded using Proposition 5 and Proposition 4, respectively. This completes the proof of (4.3). We now consider the proof of (4.4). Writing

$$\mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n\right)^2\right]$$
$$= \int \xi K_n^\ell(\mathrm{d}z_{0:n})\, \mathbb{C}_n\mathbb{S}_n(z_{0:n}, \mathrm{d}\boldsymbol{b}_n)\left(\mu(\boldsymbol{b}_n)(\mathrm{id}) - \eta_{0:n}h_n\right)^2,$$

we establish (4.4) using Corollary 3 and Proposition 5. Finally, WE consider (4.5). Using the Markov property, we obtain

$$\mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n\right)\left(\mu(\boldsymbol{\beta}_n[\ell+s])(\mathrm{id}) \ - \eta_{0:n}h_n\right)\right]$$
$$= \mathbb{E}_\xi\left[\left(\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}) - \eta_{0:n}h_n\right)\left(\mathbb{E}_{\zeta_{0:n}[\ell]}[\mu(\boldsymbol{\beta}_n[s])(\mathrm{id})] \ - \eta_{0:n}h_n\right)\right],$$

from which we may deduce (4.5) using (4.3) and (4.4).

4.2. The roll-out PPG estimator

In light of the previous results, it is natural to consider an estimator formed by an average across successive conditional PPG estimators $\{\mu(\boldsymbol{\beta}_n[\ell])\}_{\ell\in\mathbb{N}}$. To mitigate the bias, we remove a "burn-in" period, with length k_0 chosen proportionally to the mixing time of the particle Gibbs chain $\{\zeta_{0:n}[\ell]\}_{\ell\in\mathbb{N}^*}$. This yields the estimator

$$\Pi_{(k_0,k),N}(h_n) = (k-k_0)^{-1}\sum_{\ell=k_0+1}^{k}\mu(\boldsymbol{\beta}_n[\ell])(\mathrm{id}). \tag{4.6}$$

The total number of particles underlying this estimator is $C = (N-1)k$. We denote by $\upsilon = (k-k_0)/k$ the ratio of the number of particles used in the estimator to the total number of sampled particles.

As a final main result, we provide bounds on the bias and the MSE of the estimator (4.6). The proof is postponed to Section C.2.

Theorem 6. *Assume Assumption* 1. *Then, for every* $n \in \mathbb{N}$, $M \in \mathbb{N}^*$, $\xi \in \mathsf{M}_1(\mathcal{X}_{0:n})$, $\ell \in \mathbb{N}^*$, $s \in \mathbb{N}^*$, *and* $N \in \mathbb{N}^*$ *such that* $N > N_n$,

$$\left|\mathbb{E}_\xi[\Pi_{(k_0,k),N}(h_n)] - \eta_{0:n}h_n\right| \le \mathsf{c}_n^{bias}\left(\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty\right)\frac{\kappa_{N,n}^{k_0}}{N(k-k_0)(1-\kappa_{N,n})}, \tag{4.7}$$

$$\mathbb{E}_\xi\left[\left(\Pi_{(k_0,k),N}(h_n) - \eta_{0:n}h_n\right)^2\right]$$
$$\le \left(\sum_{m=0}^{n-1}\|\tilde{h}_m\|_\infty\right)^2\frac{\mathsf{c}_n^{mse} + 2\mathsf{c}_n^{cov}N^{-1/2}(1-\kappa_{N,n})^{-1}}{N(k-k_0)} \tag{4.8}$$

Setting the burn-in k_0 in the roll-out estimator is nontrivial. However, because the estimator converges for *any* choice of k_0, including the trivial choice $k_0 = 1$, we can view this algorithmic parameter as an opportunity for the user to optimize the implementation of the algorithm. For given (N,k), the choice of k_0 involves a classical trade-off between bias and variance; indeed, for fixed (N,k),

the bias upper bound (4.7) decreases with k_0 proportionally to $\kappa_{N,n}^{k_0}/(k-k_0)$ whereas the MSE upper bound (4.8) increases with k_0 proportionally to $1/(k-k_0)$. These bounds suggest that we should take $k_0 = \lceil k(1-\ell^{-1})\rceil$ if we are willing to bound the MSE increase of the roll-out estimator by a factor ℓ with respect to the PARIS. However, the bias reduction is not easily quantified, because it depends mainly on the mixing rate $\kappa_{N,n}$ of the PPG chain, and we only have access to upper bounds on this rate that are, in general, too conservative.

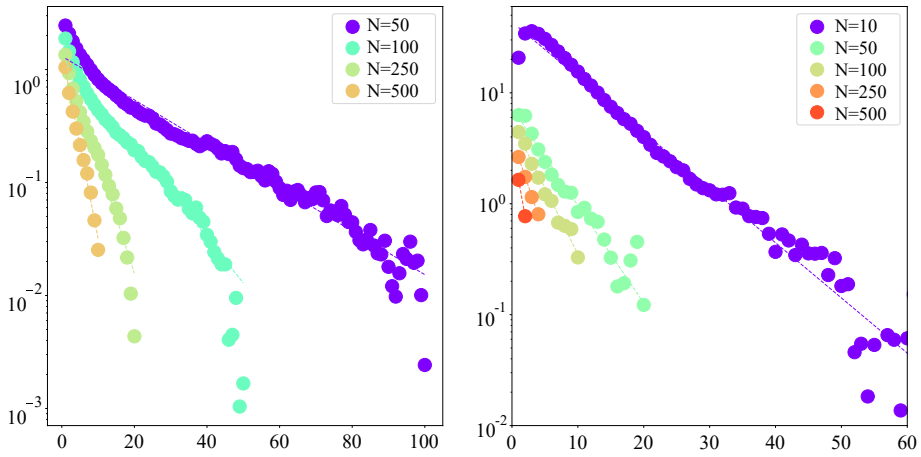
5. Numerical Results

In this section, we evaluate numerically the proposed PPG sampler in the context of general state-space HMMs. Given measurable spaces $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Z}, \mathcal{Z})$, an HMM is a bivariate (possibly inhomogeneous) Markov chain $\{(X_m, Z_m)\}_{m\in\mathbb{N}}$ taking values in the product space $(\mathsf{X}\times\mathsf{Z}, \mathcal{X}\otimes\mathcal{Z})$. In such a model, the process $\{X_n\}_{n\in\mathbb{N}}$, referred to as the *state sequence*, is assumed to be itself a (possibly inhomogeneous) Markov chain, specified by some initial distribution χ and some sequence $\{M_n\}_{n\in\mathbb{N}}$ of Markov kernels. The state sequence is latent and only partially observed through the *observation process* $\{Z_m\}_{m\in\mathbb{N}}$. Conditionally on the state sequence, the observations are assumed to be independent; furthermore, the conditional marginal distribution of each Z_m is assumed to depend only on the corresponding state X_m and to have a density $g_m(X_m, \cdot)$ with respect to some dominating measure. HMMs are used in numerous scientific and engineering disciplines; see Andrieu and Doucet (2002), Cappé, Moulines and Rydén (2005) and Chopin and Papaspiliopoulos (2020). Inference in HMMs typically involves computing conditional distributions of unobserved states, given observations. Of particular interest are the sequence of *filter distributions*, where the filter at time $m \in \mathbb{N}$, denoted as η_m, is defined as the conditional distribution of X_m, given $Z_{0:m} \coloneqq (Z_0, \dots, Z_m)$, and the *joint-smoothing distributions*, where the joint-smoothing distribution at time m, denoted as $\eta_{0:m}$, is defined as the joint conditional distribution of the states $X_{0:m} = (X_0, \dots, X_m)$, given the observations $Z_{0:m}$. Consequently, η_m is the marginal of $\eta_{0:m}$ with respect to the last state X_m. Given a sequence $\{z_m\}_{m\in\mathbb{N}}$ of fixed observations, $\{\eta_{0:m}\}_{m\in\mathbb{N}}$ forms a Feynman–Kac model (see Section 1), with Markov kernels $\{M_m\}_{m\in\mathbb{N}}$ and potential functions $g_m \coloneqq g(\cdot, z_m)$, for $m \in \mathbb{N}$, on X.

We now evaluate the proposed algorithm numerically for two HMMs: (i) a linear Gaussian state-space model (for which the filter and the joint-smoothing distribution flows are available in a closed form), and (ii) the stochastic volatility model proposed in Hull and White (1987). The PPG algorithm used in this section is given in Algorithm 3 (in Section B).

Linear Gaussian state-space model (LGSSM). We first consider an LGSSM

$$X_{m+1} = AX_m + Q\epsilon_{m+1}, \quad Z_m = BX_m + R\zeta_m, \quad m \in \mathbb{N}, \tag{5.1}$$



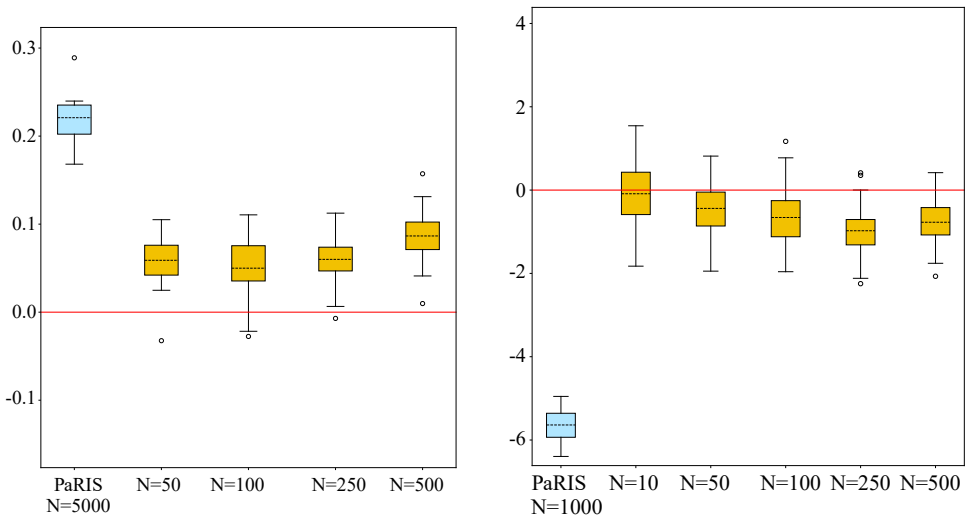
(a) LGSMM (b) StoVol

Figure 1. Output of the PPG roll-out estimator for the LGSSM (left panel) and the StoVol model (right panel). The curves describe the evolution of the bias with increasing k for different batch sizes N.

where $\{\epsilon_m\}_{m\in\mathbb{N}^*}$ and $\{\zeta_m\}_{m\in\mathbb{N}}$ are sequences of independent standard normally distributed random variables. The matrices A, Q, B, and R are assumed to be known 5×5 matrices (see Appendix A.1 for the precise values). In this framework, we aim to compute the expectation of the *one-lag state covariance* $h_n(x_{0:n}) \coloneqq \sum_{m=0}^{n-1} x_m x_{m+1}^\intercal$ under the joint-smoothing distribution $\eta_{0:n}$ for observations generated by simulation under the given parameters with $n = 10^3$. In the LGSSM case, the *disturbance smoother* (see Cappé, Moulines and Rydén (2005, Algo. 5.2.15)) provides the exact values of $\eta_{0:n}h_n$, which allows us to assess numerically the bias of the PARIS and PPG estimators.

In this setting, we calculate the bias for batch sizes $N \in \{10, 25, 50, 100, 500\}$ and an increasing number k of iterations by averaging the PPG estimator over 10^4 independent runs. Figure 1a shows the bias of the PPG estimates of the first diagonal entry of the one-lag covariance. For each batch size N, we estimate and display the regression function $k \mapsto \mathrm{e}^{ak+b}$ to illustrate the exponential decrease of the PPG bias, which is consistent with Theorem 2.

Figure 2a displays, for a given budget $C = 5 \times 10^3$, the bias of the estimates of $\eta_{0:n}h_n$ using the PARIS and the PPG for different batch sizes N and different numbers $k = C/N$ of iterations and burn-in periods $k_0 = \lfloor k/2 \rfloor$. The red line corresponds to zero (no bias), and the empirical means are given by black-dashed lines. An extended comparison comprising different choices of k_0 and different budgets C is provided in Section A. In order to estimate the bias for each algorithmic configuration, we average 10^3 independent replications of the corresponding estimator. Moreover, to assess the precision of the resulting bias



(a) LGSSM (b) StoVol

Figure 2. PARIS and PPG bias dispersions for the LGSSM and StoVol model as a function of the mini-batch size N for fixed computational budgets $C = Nk$ of 5×10^3 (LGSSM) and 10^3 (StoVol model) and with $k_0 = \lfloor 2^{-1}k \rfloor$ burn-in steps.

estimator, we repeat this procedure 10^2 times, and present the bias estimates in a box plot. This enables us to form an idea of whether the PPG provides a statistically significant improvement in terms of bias. In this example, whatever the choice of the batch size is, the PPG bias is significantly reduced compared with the bias of the PARIS estimator. We further observe that a larger k leads to smaller bias.

Stochastic volatility (StoVol). As a second example, consider the stochastic volatility model

$$X_{m+1} = \phi X_m + \sigma_\epsilon \epsilon_{m+1}, \quad Z_m = \beta \exp\left(\frac{X_m}{2}\right)\zeta_m, \quad m \in \mathbb{N}, \tag{5.2}$$

where $\{\epsilon_m\}_{m\in\mathbb{N}^*}$ and $\{\zeta_m\}_{m\in\mathbb{N}}$ are as in the previous example, and the model parameters ϕ, β, and σ_ϵ are set to 0.975, 0.63, and 0.16, respectively. The reference value is calculated by running the PARIS with 5×10^4 particles. In this setting, we repeated the experiments of the previous example for the same additive functional and number $n = 10^3$ of observations, produced by simulation under the parameters above. The computational budget was set to $C = 10^3$. As in the LGSSM example, the bias decay with respect to the iteration index k is displayed in Figure 1b, and the comparison with the PARIS is shown in Figure 2b. The comments from the previous example apply to this StoVol model context as well. More in-depth numerical assessments of the proposed PPG estimator are

found in Section A.2. In particular, in Section A.2.1, we compare our estimator with the Rhee–Glynn-type estimator with ancestor sampling proposed by Jacob, Lindsten and Schön (2020), showing that the variance of the latter is significantly larger than that of the PPG for a given computational effort.

6. Proofs

6.1. Proof of Proposition 1

Using the identity

$$\eta_0 Q_0 \cdots Q_{n-1} \mathbb{1}_{\mathsf{X}_n} = \prod_{m=0}^{n-1} \eta_m Q_m \mathbb{1}_{\mathsf{X}_{m+1}}$$

and that each kernel Q_m has a transition density, write, for $h \in \mathsf{F}(\mathcal{X}_{0:n})$,

$$\begin{aligned}
\eta_{0:n} h &= \int \cdots \int h(x_{0:n})\, \eta_0(\mathrm{d}x_0) \prod_{m=0}^{n-1} \left(\frac{\eta_m[q_m(\cdot, x_{m+1})]\, \lambda_{m+1}(\mathrm{d}x_{m+1})}{\eta_m Q_m \mathbb{1}_{\mathsf{X}_{m+1}}} \right) \\
&\quad \left(\frac{q_m(x_m, x_{m+1})}{\eta_m[q_m(\cdot, x_{m+1})]} \right) \\
&= \int \cdots \int h(x_{0:n})\, \eta_n(\mathrm{d}x_n) \prod_{m=0}^{n-1} \frac{\eta_m(\mathrm{d}x_m)\, q_m(x_m, x_{m+1})}{\eta_m[q_m(\cdot, x_{m+1})]} \qquad (6.1) \\
&= \left(\overleftarrow{Q}_{0,\eta_0} \otimes \cdots \otimes \overleftarrow{Q}_{n-1,\eta_{n-1}} \otimes \eta_n \right) h,
\end{aligned}$$

which establishes the proof.

6.2. Proof of Theorem 1

Lemma 1. *For all $n \in \mathbb{N}$, $\boldsymbol{x}_n \in \mathbf{X}_n$, and $h \in \mathsf{F}(\boldsymbol{\mathcal{X}}_{n+1} \otimes \mathcal{X}_{n+1})$,*

$$\begin{aligned}
&\iint h(\boldsymbol{x}_{n+1}, z_{n+1})\, \boldsymbol{Q}_n(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1})\, \mu(\boldsymbol{x}_{n+1})(\mathrm{d}z_{n+1}) \\
&= \iint h(\boldsymbol{x}_{n+1}, z_{n+1})\, \mu(\boldsymbol{x}_n) Q_n(\mathrm{d}z_{n+1})\, \boldsymbol{M}_n\langle z_{n+1}\rangle(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1}). \qquad (6.2)
\end{aligned}$$

In addition, for all $h \in \mathsf{F}(\boldsymbol{\mathcal{X}}_0 \otimes \mathcal{X}_0)$,

$$\iint h(\boldsymbol{x}_0, z_0)\, \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0)\, \mu(\boldsymbol{x}_0)(\mathrm{d}z_0) = \iint h(\boldsymbol{x}_0, z_0)\, \boldsymbol{\eta}_0\langle z_0\rangle(\mathrm{d}\boldsymbol{x}_0)\, \eta_0(\mathrm{d}z_0). \qquad (6.3)$$

Proof. Because $\mu(\boldsymbol{x}_n)\, Q_n(\mathrm{d}z_{n+1}) = \boldsymbol{g}_n(\boldsymbol{x}_n)\, \Phi_n(\mu(\boldsymbol{x}_n))(\mathrm{d}z_{n+1})$, we may rewrite the right-hand side of (6.2) as

$$\iint h(\boldsymbol{x}_{n+1}, z_{n+1})\, \mu(\boldsymbol{x}_n) Q_n(\mathrm{d}z_{n+1})\, \boldsymbol{M}_n\langle z_{n+1}\rangle(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1})$$

$$
\begin{aligned}
&= \boldsymbol{g}_n(\boldsymbol{x}_n)\frac{1}{N}\sum_{i=0}^{N-1}\iint h(\boldsymbol{x}_{n+1}, z_{n+1})\,\Phi_n(\mu(\boldsymbol{x}_n))(\mathrm{d}z_{n+1}) \\
&\quad\times\left(\Phi_n(\mu(\boldsymbol{x}_n))^{\otimes i}\otimes\delta_{z_{n+1}}\otimes\Phi_n(\mu(\boldsymbol{x}_n))^{\otimes(N-i-1)}\right)(\mathrm{d}\boldsymbol{x}_{n+1}) \\
&= \boldsymbol{g}_n(\boldsymbol{x}_n)\frac{1}{N}\sum_{i=1}^{N}\int\cdots\int h((x_{n+1}^1,\ldots,x_{n+1}^{i-1}, z_{n+1}, x_{n+1}^{i+1},\ldots,x_{n+1}^N), z_{n+1}) \\
&\quad\times\Phi_n(\mu(\boldsymbol{x}_n))(\mathrm{d}z_{n+1})\prod_{\ell\neq i}\Phi_n(\mu(\boldsymbol{x}_n))(\mathrm{d}x_{n+1}^{\ell}) \\
&= \boldsymbol{g}_n(\boldsymbol{x}_n)\frac{1}{N}\sum_{i=1}^{N}\int h(\boldsymbol{x}_{n+1}, x_{n+1}^i)\,\boldsymbol{M}_n(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1}).
\end{aligned}
$$

On the other hand, note that the left-hand side of (6.2) can be expressed as

$$
\begin{aligned}
&\iint h(\boldsymbol{x}_{n+1}, z_{n+1})\,\boldsymbol{Q}_n(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1})\,\mu(\boldsymbol{x}_{n+1})(\mathrm{d}z_{n+1}) \\
&= \boldsymbol{g}_n(\boldsymbol{x}_n)\frac{1}{N}\sum_{i=1}^{N}\int h(\boldsymbol{x}_{n+1}, x_{n+1}^i)\,\boldsymbol{M}_n(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1}),
\end{aligned}
\tag{6.4}
$$

which establishes the identity. The identity (6.3) is established along similar lines.

We establish Theorem 1 by induction. Thus, assume that the claim holds for n, and show that for all $h \in \mathsf{F}(\boldsymbol{\mathcal{X}}_{0:n+1}\otimes\mathcal{X}_{0:n+1})$,

$$
\begin{aligned}
&\iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\boldsymbol{\gamma}_{0:n+1}(\mathrm{d}\boldsymbol{x}_{0:n+1})\,\mathbb{B}_{n+1}(\boldsymbol{x}_{0:n+1}, \mathrm{d}z_{0:n+1}) \\
&= \iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\gamma_{0:n+1}(\mathrm{d}z_{0:n+1})\,\mathbb{C}_{n+1}(z_{0:n+1}, \mathrm{d}\boldsymbol{x}_{0:n+1}).
\end{aligned}
\tag{6.5}
$$

To prove this, we process, using definition (2.4), the left-hand side of (6.5) according to

$$
\begin{aligned}
&\iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\boldsymbol{\gamma}_{0:n+1}(\mathrm{d}\boldsymbol{x}_{0:n+1})\,\mathbb{B}_{n+1}(\boldsymbol{x}_{0:n+1}, \mathrm{d}z_{0:n+1}) \\
&= \iint \boldsymbol{\gamma}_{0:n}(\mathrm{d}\boldsymbol{x}_{0:n})\,\mathbb{B}_n(\boldsymbol{x}_{0:n}, \mathrm{d}z_{0:n}) \\
&\quad\times\iint \bar{h}(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\boldsymbol{Q}_n(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1})\,\mu(\boldsymbol{x}_{n+1})(\mathrm{d}z_{n+1}),
\end{aligned}
\tag{6.6}
$$

where we define the function

$$
\bar{h}(\boldsymbol{x}_{0:n+1}, z_{0:n+1}) \coloneqq \frac{q_n(z_n, z_{n+1})h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})}{\mu(\boldsymbol{x}_n)[q_n(\cdot, z_{n+1})]}.
$$

Now, applying Lemma 1 to the inner integral and using

$$\mu(\boldsymbol{x}_n)Q_n(\mathrm{d}z_{n+1}) = \mu(\boldsymbol{x}_n)[q_n(\cdot, z_{n+1})]\,\lambda_{n+1}(\mathrm{d}z_{n+1})$$

yields, for every $\boldsymbol{x}_{0:n}$ and $z_{0:n}$,

$$\begin{aligned}
&\iint \bar{h}(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\boldsymbol{Q}_n(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1})\,\mu(\boldsymbol{x}_{n+1})(\mathrm{d}z_{n+1}) \\
&= \iint \bar{h}(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\mu(\boldsymbol{x}_n)Q_n(\mathrm{d}z_{n+1})\,\boldsymbol{M}_n\langle z_{n+1}\rangle(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1}) \\
&= \iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,Q_n(z_n, \mathrm{d}z_{n+1})\,\boldsymbol{M}_n\langle z_{n+1}\rangle(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1}).
\end{aligned}$$

Inserting the previous identity into (6.6) and using the induction hypothesis yields

$$\begin{aligned}
&\iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\boldsymbol{\gamma}_{0:n+1}(\mathrm{d}\boldsymbol{x}_{0:n+1})\,\mathbb{B}_{n+1}(\boldsymbol{x}_{0:n+1}, \mathrm{d}z_{0:n+1}) \\
&= \iint \gamma_{0:n}(\mathrm{d}z_{0:n})\,\mathbb{C}_n(z_{0:n}, \mathrm{d}\boldsymbol{x}_{0:n}) \\
&\quad \times \iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,Q_n(z_n, \mathrm{d}z_{n+1})\,\boldsymbol{M}_n\langle z_{n+1}\rangle(\boldsymbol{x}_n, \mathrm{d}\boldsymbol{x}_{n+1}) \\
&= \iint h(\boldsymbol{x}_{0:n+1}, z_{0:n+1})\,\gamma_{0:n+1}(\mathrm{d}z_{0:n+1})\,\mathbb{C}_{n+1}(z_{0:n+1}, \mathrm{d}\boldsymbol{x}_{0:n+1}),
\end{aligned}$$

which establishes (6.5).

6.3. Proof of Theorem 3

First, define, for $m \in \mathbb{N}$,

$$\boldsymbol{P}_{2=}\langle m\rangle : \mathbf{Y}_m \times \boldsymbol{\mathcal{Y}}_{m+1} \ni (\boldsymbol{y}_m, A) \mapsto \int \boldsymbol{M}_m(\boldsymbol{x}_{m|m}, \mathrm{d}\boldsymbol{x}_{m+1})\,\boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, A). \tag{6.7}$$

For any given initial distribution $\boldsymbol{\psi}_0 \in \mathsf{M}_1(\boldsymbol{\mathcal{Y}}_0)$, let $\mathbb{P}^{\boldsymbol{P}}_{\boldsymbol{\psi}_0}$ be the distribution of the canonical Markov chain induced by the Markov kernels $\{\boldsymbol{P}_m\}_{m\in\mathbb{N}}$ and the initial distribution $\boldsymbol{\psi}_0$. With a slight abuse of notation we write, for $\boldsymbol{\eta}_0 \in \mathsf{M}_1(\boldsymbol{\mathcal{X}}_0)$, $\mathbb{P}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}$ instead of $\mathbb{P}^{\boldsymbol{P}}_{\boldsymbol{\psi}_0[\boldsymbol{\eta}_0]}$, where we define the extension $\boldsymbol{\psi}_0[\boldsymbol{\eta}_0](A) = \int \mathbb{1}_A(\mathbf{J}\boldsymbol{x}_0)\,\boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0)$, for $A \in \boldsymbol{\mathcal{Y}}_0$. We preface the proof of Theorem 3 with some technical lemmas and a proposition.

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

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

Proof. Pick arbitrary $\varphi \in \mathsf{F}(\mathcal{X}_{n:n+1})$ and, from definition (2.3) and that Q_n has a transition density, write

$$\begin{aligned}
&\iint \varphi(x_{n:n+1})\,\gamma_n(\mathrm{d}x_n)\,Q_n(x_n,\mathrm{d}x_{n+1}) \\
&= \iint \varphi(x_{n:n+1})\gamma_n[q_n(\cdot,x_{n+1})]\,\lambda_{n+1}(\mathrm{d}x_{n+1})\,\frac{\gamma_n(\mathrm{d}x_n)q_n(x_n,x_{n+1})}{\gamma_n[q_n(\cdot,x_{n+1})]} \\
&= \iint \varphi(x_{n:n+1})\,\gamma_{n+1}(\mathrm{d}x_{n+1})\,\overleftarrow{Q}_{n,\eta_n}(x_{n+1},\mathrm{d}x_n). \qquad (6.8)
\end{aligned}$$

Now, by (2.10), it holds that

$$\begin{aligned}
&B_{n+1}h_{n+1}(x_{n+1}) \\
&= \int \overleftarrow{Q}_{n,\eta_n}(x_{n+1},\mathrm{d}x_n)\left(\tilde{h}_n(x_{n:n+1}) + \int h_n(x_{0:n})\,B_n(x_n,\mathrm{d}x_{0:n-1})\right);
\end{aligned}$$

therefore, by applying (6.8) with

$$\varphi(x_{n:n+1}) \coloneqq f_{n+1}(x_{n+1})\left(\tilde{h}_n(x_{n:n+1}) + \int h_n(x_{0:n})\,B_n(x_n,\mathrm{d}x_{0:n-1})\right),$$

we obtain that

$$\begin{aligned}
\gamma_{n+1}(f_{n+1}B_{n+1}h_{n+1}) &= \iint \varphi(x_{n:n+1})\,\gamma_{n+1}(\mathrm{d}x_{n+1})\,\overleftarrow{Q}_{n,\eta_n}(x_{n+1},\mathrm{d}x_n) \\
&= \iint \varphi(x_{n:n+1})\,\gamma_n(\mathrm{d}x_n)\,Q_n(x_n,\mathrm{d}x_{n+1}) \\
&= \gamma_n(Q_n f_{n+1}B_n h_n + Q_n\tilde{h}_n f_{n+1}).
\end{aligned}$$

Now, the proof is concluded by noting that because $\gamma_{n+1} = \gamma_n Q_n$, $\gamma_{n+1}\tilde{f}_{n+1} = \gamma_n Q_n \tilde{f}_{n+1}$.

Lemma 3. *For every $n \in \mathbb{N}^*$, $h_n \in \mathsf{F}(\mathcal{Y}_n)$, and $\boldsymbol{\eta}_0 \in \mathsf{M}_1(\boldsymbol{\mathcal{X}}_0)$, it holds that*

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[h_n(\boldsymbol{\upsilon}_n) \mid \boldsymbol{\xi}_{0|0},\dots,\boldsymbol{\xi}_{n|n}] = \mathbb{S}_n h_n(\boldsymbol{\xi}_{0|0},\dots,\boldsymbol{\xi}_{n|n}), \quad \mathbb{P}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\text{-a.s.}$$

Proof. Pick arbitrary $v_n \in \mathsf{F}(\mathcal{X}_{0:n})$. We show that

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[v_n(\boldsymbol{\xi}_{0|0},\dots,\boldsymbol{\xi}_{n|n})h_n(\boldsymbol{\upsilon}_n)] = \mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[v_n(\boldsymbol{\xi}_{0|0},\dots,\boldsymbol{\xi}_{n|n})\mathbb{S}_n h_n(\boldsymbol{\xi}_{0|0},\dots,\boldsymbol{\xi}_{n|n})], \quad (6.9)$$

from which the claim follows. Using definition (6.7), the left-hand side of the previous identity may be rewritten as

$$\begin{aligned}
&\int\cdots\int \boldsymbol{\psi}_0[\boldsymbol{\eta}_0](\mathrm{d}\boldsymbol{y}_0)\prod_{m=0}^{n-1}\boldsymbol{P}_m(\boldsymbol{y}_m,\mathrm{d}\boldsymbol{y}_{m+1})\,h_n(\boldsymbol{y}_n)v_n(\boldsymbol{x}_{0|0},\dots,\boldsymbol{x}_{n|n}) \\
&= \int\cdots\int \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_{0|0})\prod_{m=0}^{n-1}\boldsymbol{M}_m(\boldsymbol{x}_{m|m},\mathrm{d}\boldsymbol{x}_{m+1})\,\boldsymbol{S}_0(\mathbf{J}\boldsymbol{x}_{0|0},\boldsymbol{x}_1,\mathrm{d}\boldsymbol{y}_1)
\end{aligned}$$

$$\times \prod_{m=0}^{n-1} \boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, \mathrm{d}\boldsymbol{y}_{m+1})\, h_n(\boldsymbol{y}_n) v_n(\boldsymbol{x}_{0|0}, \dots, \boldsymbol{x}_{n|n})$$
$$= \int \cdots \int \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0) \prod_{m=0}^{n-1} \boldsymbol{M}_m(\boldsymbol{x}_m, \mathrm{d}\boldsymbol{x}_{m+1})\, \boldsymbol{S}_0(\mathbf{J}\boldsymbol{x}_0, \boldsymbol{x}_1, \mathrm{d}\boldsymbol{y}_1)$$
$$\times \prod_{m=0}^{n-1} \boldsymbol{S}_m(\boldsymbol{y}_m, \boldsymbol{x}_{m+1}, \mathrm{d}\boldsymbol{y}_{m+1})\, h_n(\boldsymbol{y}_n) v_n(\boldsymbol{x}_0, \dots, \boldsymbol{x}_n).$$

Thus, we conclude the proof by using the definition (3.2) of $\mathbb{S}_n$, together with Fubini's theorem.

Lemma 4. *For every $n \in \mathbb{N}^*$ and $h_n \in \mathsf{F}(\mathcal{Y}_n)$, it holds that*

$$\mathbb{E}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) h_n(\boldsymbol{\upsilon}_n)\right] = \int \boldsymbol{\gamma}_{0:n} \mathbb{S}_n(\mathrm{d}\boldsymbol{y}_n)\, h_n(\boldsymbol{y}_n).$$

Proof. The claim of the lemma is a direct implication of Lemma 3; indeed, by applying the tower property and the latter, we obtain

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) h_n(\boldsymbol{\upsilon}_n)\right]$$
$$= \mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \mathbb{S}_n h_n(\boldsymbol{\xi}_{0|0}, \dots, \boldsymbol{\xi}_{n|n})\right]$$
$$= \int \cdots \int \boldsymbol{\eta}_0(\mathrm{d}\boldsymbol{x}_0) \prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{x}_m)\, \boldsymbol{M}_m(\boldsymbol{x}_m, \mathrm{d}\boldsymbol{x}_{m+1})\, \mathbb{S}_n h_n(\boldsymbol{x}_{0:n})$$
$$= \int \boldsymbol{\gamma}_{0:n} \mathbb{S}_n(\mathrm{d}\boldsymbol{y}_n)\, h_n(\boldsymbol{y}_n).$$

Proposition 6. *For all $n \in \mathbb{N}^*$, $(N, M) \in (\mathbb{N}^*)^2$, and $(f_n, \tilde{f}_n) \in \mathsf{F}(\mathcal{X}_n)^2$,*

$$\int \boldsymbol{\gamma}_{0:n} \mathbb{S}_n(\mathrm{d}\boldsymbol{y}_n) \left(\frac{1}{N} \sum_{i=1}^{N} \{b_n^i f_n(x_{n|n}^i) + \tilde{f}_n(x_{n|n}^i)\}\right) = \gamma_n(f_n B_n h_n + \tilde{f}_n).$$

Proof. Applying Lemma 4 yields

$$\int \boldsymbol{\gamma}_{0:n} \mathbb{S}_n(\mathrm{d}\boldsymbol{y}_n) \left(\frac{1}{N} \sum_{i=1}^{N} \{b_n^i f_n(x_{n|n}^i) + \tilde{f}_n(x_{n|n}^i)\}\right)$$
$$= \mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \frac{1}{N} \sum_{i=1}^{N} \{\beta_n^i f_n(\xi_{n|n}^i) + \tilde{f}_n(\xi_{n|n}^i)\}\right]. \tag{6.10}$$

In the following, we repeatedly use the following filtrations. Let $\tilde{\mathcal{F}}_n \coloneqq \sigma(\{\boldsymbol{\upsilon}_m\}_{m=0}^n)$ be the σ-field generated by the output of the `PARIS` (Algorithm

1) during the first n iterations. In addition, let $\mathcal{F}_n := \tilde{\mathcal{F}}_{n-1} \vee \sigma(\boldsymbol{\xi}_{n|n})$.

We proceed by induction. Thus, assume that the statement of the proposition holds for a given $n \in \mathbb{N}^*$, and consider, for arbitrarily chosen $(f_{n+1}, \tilde{f}_{n+1}) \in \mathsf{F}(\mathcal{X}_{n+1})^2$,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \frac{1}{N}\sum_{i=1}^{N}\{\beta^i_{n+1} f_{n+1}(\xi^i_{n+1|n+1}) + \tilde{f}_{n+1}(\xi^i_{n+1|n+1})\}\tilde{\mathcal{F}}_n\right] \\
&= \left(\prod_{m=0}^{n} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\beta^1_{n+1} f_{n+1}(\xi^1_{n+1|n+1}) + \tilde{f}_{n+1}(\xi^1_{n+1|n+1})\tilde{\mathcal{F}}_n\right],
\end{aligned}$$

where we use that the variables $\{\beta^i_{n+1} f_{n+1}(\xi^i_{n+1|n+1}) + \tilde{f}_{n+1}(\xi^i_{n+1|n+1})\}_{i=1}^N$ are conditionally independent and identically distributed (i.i.d.) given $\tilde{\mathcal{F}}_n$. Note that, by symmetry,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[\beta^1_{n+1} \mid \mathcal{F}_{n+1}] = \int \boldsymbol{S}_n(\boldsymbol{\upsilon}_n, \boldsymbol{\xi}_{n+1|n+1}, \mathrm{d}\boldsymbol{y}_{n+1})\, b^1_{n+1} \\
&= \int \cdots \int \left(\prod_{j=1}^{M}\sum_{\ell=1}^{N} \frac{q_n(\xi^\ell_{n|n}, \xi^1_{n+1|n+1})}{\sum_{\ell'=1}^{N} q_n(\xi^{\ell'}_{n|n}, \xi^1_{n+1|n+1})} \delta_{(\xi^\ell_{n|n}, \beta^\ell_n)}(\mathrm{d}\tilde{x}^{1,j}_n, \mathrm{d}\tilde{b}^{1,j}_n)\right) \\
&\quad \times \frac{1}{M}\sum_{j=1}^{M}\left(\tilde{b}^{1,j}_n + \tilde{h}_n(\tilde{x}^{1,j}_n, \xi^1_{n+1|n+1})\right) \\
&= \sum_{\ell=1}^{N} \frac{q_n(\xi^\ell_{n|n}, \xi^1_{n+1|n+1})}{\sum_{\ell'=1}^{N} q_n(\xi^{\ell'}_{n|n}, \xi^1_{n+1|n+1})}\left(\beta^\ell_n + \tilde{h}_n(\xi^\ell_{n|n}, \xi^1_{n+1|n+1})\right). \qquad (6.11)
\end{aligned}$$

Thus, using the tower property,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[\beta^1_{n+1} f_{n+1}(\xi^1_{n+1|n+1}) \mid \tilde{\mathcal{F}}_n] = \\
&\int \Phi_n(\mu(\boldsymbol{\xi}_{n|n}))(dx_{n+1})\, f_{n+1}(x_{n+1}) \sum_{\ell=1}^{N} \frac{q_n(\xi^\ell_{n|n}, x_{n+1})}{\sum_{\ell'=1}^{N} q_n(\xi^{\ell'}_{n|n}, x_{n+1})}\Big(\beta^\ell_n + \tilde{h}_n(\xi^\ell_{n|n}, x_{n+1})\Big),
\end{aligned}$$

and, consequently, using definition (2.1),

$$\begin{aligned}
&\left(\prod_{m=0}^{n} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}[\beta^1_{n+1} f_{n+1}(\xi^1_{n+1|n+1}) \mid \tilde{\mathcal{F}}_n] \\
&= \left(\prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \int \frac{1}{N}\sum_{i=1}^{N} q_n(\xi^i_{n|n}, x_{n+1}) \qquad (6.12) \\
&\quad \times f_{n+1}(x_{n+1}) \sum_{\ell=1}^{N} \frac{q_n(\xi^\ell_{n|n}, x_{n+1})}{\sum_{\ell'=1}^{N} q_n(\xi^{\ell'}_{n|n}, x_{n+1})}\left(\beta^\ell_n + \tilde{h}_n(\xi^\ell_{n|n}, x_{n+1})\right) \lambda_{n+1}(\mathrm{d}x_{n+1}) \\
&= \left(\prod_{m=0}^{n-1} \boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right) \frac{1}{N}\sum_{\ell=1}^{N}\left(\beta^\ell_n Q_n f_{n+1}(\xi^\ell_{n|n}) + Q_n(\tilde{h}_n f_{n+1})(\xi^\ell_{n|n})\right).
\end{aligned}$$

Thus, applying the induction hypothesis,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\beta^i_{n+1}f_{n+1}(\xi^i_{n+1|n+1})\right]\\
&=\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n-1}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{\ell=1}^{N}\left(\beta^\ell_n Q_n f_{n+1}(\xi^\ell_{n|n})+Q_n(\tilde{h}_n f_{n+1})(\xi^\ell_{n|n})\right)\right]\\
&=\gamma_n\left(Q_n f_{n+1}B_n h_n+Q_n(\tilde{h}_n f_{n+1})\right). \qquad (6.13)
\end{aligned}$$

In the same manner, it can be shown that

$$\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\tilde{f}_{n+1}(\xi^i_{n+1|n+1})\right]=\gamma_n Q_n\tilde{f}_{n+1}. \qquad (6.14)$$

Now, by (6.13–6.14) and Lemma 2,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{P}}_{\boldsymbol{\eta}_0}\left[\left(\prod_{m=0}^{n}\boldsymbol{g}_m(\boldsymbol{\xi}_{m|m})\right)\frac{1}{N}\sum_{i=1}^{N}\{\beta^i_{n+1}f_{n+1}(\xi^i_{n+1|n+1})+\tilde{f}_{n+1}(\xi^i_{n+1|n+1})\}\right]\\
&=\gamma_n\left(Q_n f_{n+1}B_n h_n+Q_n(\tilde{h}_n f_{n+1}+Q_n\tilde{f}_{n+1})\right)\\
&=\gamma_{n+1}(f_{n+1}B_{n+1}h_{n+1}+\tilde{f}_{n+1}),
\end{aligned}$$

which shows that the claim of the proposition holds at time $n+1$.

It remains to check the base case $n=0$, which holds trivially, because $\boldsymbol{\beta}_0=\mathbf{0}$ and $B_0h_0=0$ by convention, and the initial particles $\boldsymbol{\xi}_{0|0}$ are drawn from $\boldsymbol{\eta}_0$. This completes the proof.

Proof of Theorem 3. The identity $\int\boldsymbol{\eta}_{0:n}(\mathrm{d}\boldsymbol{x}_{0:n})\,\mathbb{S}_n(\boldsymbol{x}_{0:n},\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id}) = \eta_{0:n}h_n$ follows immediately by letting $f_n\equiv 1$ and $\tilde{f}_n\equiv 0$ in Proposition 6, and using that $\boldsymbol{\gamma}_{0:n}(\mathbf{X}_{0:n})=\gamma_{0:n}(\mathsf{X}_{0:n})$. Moreover, applying Theorem 1 yields

$$\begin{aligned}
&\int\eta_{0:n}\mathbb{C}_n\mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id})\\
&=\iint\eta_{0:n}(\mathrm{d}z_{0:n})\,\mathbb{C}_n(z_{0:n},\mathrm{d}\boldsymbol{x}_{0:n})\int\mathbb{S}_n(\boldsymbol{x}_{0:n},\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id})\\
&=\iint\boldsymbol{\eta}0:n(\mathrm{d}\boldsymbol{x}_{0:n})\,\mathbb{B}_n(\boldsymbol{x}_{0:n},\mathrm{d}z_{0:n})\int\mathbb{S}_n(\boldsymbol{x}_{0:n},\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id})\\
&=\int\boldsymbol{\eta}0:n\mathbb{S}_n(\mathrm{d}\boldsymbol{b}_n)\,\mu(\boldsymbol{b}_n)(\mathrm{id}).
\end{aligned}$$

Finally, the first identity holds because K_n leaves $\eta_{0:n}$ invariant.

Supplementary Material

The supplementary material contains proofs for the technical propositions, lemmas and theorems as well as additional numerical investigations of different aspects of the `PPG` algorithm.

Acknowledgments

The work of J. Olsson was supported by the Swedish Research Council, Grant 2018-05230. The work of E. Moulines was supported by ANR CHIA 002 – SCAI. The work of G. Cardoso was supported by the Investment of the Future Grant, ANR-10-IAHU-04, from the government of France through the Agence National de la Recherche.

References

Andrieu, C. and Doucet, A. (2002). Particle filtering for partially observed Gaussian state space models. *J. Roy. Statist. Soc. B* **64**, 827–836.

Andrieu, C. and Doucet, A. (2003). Online Expectation–Maximization type algorithms for parameter estimation in general state space models. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 69–72.

Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. B* **72**, 269–342.

Cappé, O. (2001). Recursive computation of smoothed functionals of hidden Markovian processes using a particle approximation. *Monte Carlo Methods Appl.* **7**, 81–92.

Cappé, O. (2011). Online EM algorithm for hidden Markov models. *J. Comput. Graph. Statist.* **20**, 728–749.

Cappé, O., Godsill, S. J. and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *IEEE Proceedings* **95**, 899–924.

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer.

Chopin, N. and Papaspiliopoulos, O. (2020). *An Introduction to Sequential Monte Carlo*. Springer.

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

Del Moral, P. (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer.

Del Moral, P. (2013). *Mean Field Simulation for Monte Carlo Integration*. CRC Press.

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

Del Moral, P. and Jasra, A. (2018). A sharp first order analysis of Feynman–Kac particle models, part II: Particle Gibbs samplers. *Stoch. Proc. Appl.* **128**, 354–371.

Del Moral, P., Kohn, R. and Patras, F. (2016). On particle Gibbs samplers. *Ann. Inst. H. Poincaré Probab. Statist.* **52**, 1687–1733.

Douc, R., Garivier, A., Moulines, E. and Olsson, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21**, 1201–2145.

Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36**, 2344–2376.

Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov Chains*. Springer.

Gloaguen, P., Le Corff, S. and Olsson, J. (2022). A pseudo-marginal sequential Monte Carlo online smoothing algorithm. *Bernoulli* **28**, 2606–2633.

Godsill, S. J., Doucet, A. and West, M. (2004). Monte Carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.* **50**, 438–449.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42**, 281–300.

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

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

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

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

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.* **94**, 590–599.

Poyiadjis, G., Doucet, A. and Singh, S. S. (2005). Particle methods for optimal filter derivative: application to parameter estimation. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* v/925–v/928.

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.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.

Whiteley, N. (2010). Discussion on particle Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B* **72**, 306–307.

Gabriel Cardoso

Ecole Polytechnique - CMAP, Palaiseau 91120, France.

E-mail: gvictorinocardoso@gmail.com

Eric Moulines

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

E-mail: eric.moulines@polytechnique.edu

Jimmy Olsson

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

E-mail: jimmyol@kth.se

(Received June 2022; accepted March 2023)