

BIAS OF PARTICLE APPROXIMATIONS TO OPTIMAL FILTER DERIVATIVE

VLADISLAV Z. B. TADIĆ * AND ARNAUD DOUCET †

Abstract. In many applications, a state-space model depends on a parameter which needs to be inferred from data in an online manner. In the maximum likelihood approach, this can be achieved using stochastic gradient search, where the underlying gradient estimation is based on the optimal filter and the optimal filter derivative. However, the optimal filter and its derivative are not analytically tractable for a non-linear state-space model and need to be approximated numerically. In [22], a particle approximation to this derivative has been proposed, while the corresponding central limit theorem and L_p error bounds have been established in [11]. We derive here bounds on the bias of this particle approximation. Under mixing conditions, these bounds are uniform in time and inversely proportional to the number of particles.

Key words. Particle Methods, Bias, Optimal Filter, Optimal Filter Derivative, Non-Linear State-Space Models.

AMS subject classifications. Primary 93E11; Secondary 62M20, 65C05.

1. Introduction. State-space models, also known as continuous-state hidden Markov models, are a class of stochastic processes used to model complex time-series data and stochastic dynamical systems. A state-space model can be described as a latent discrete-time Markov process observed only through noisy measurements of its states. In this context, one of the most important problems is the optimal estimation of the present state given the noisy observations of the present and past states. This problem is known as optimal filtering. For non-linear state-space models, the optimal filter does not typically admit a closed-form expression and needs to be approximated numerically. Numerous numerical methods for optimal filtering have been proposed and studied in the literature — see e.g., [4] for a recent overview. Among them, particle methods (also known as sequential Monte Carlo methods) have gained significant attention. Their convergence properties have been thoroughly investigated in a number of papers and books — see, e.g., [3], [4], [6], [7], [14], [15].

In many scenarios of practical interest, a state-space model depends on a parameter whose value needs to be estimated given a set of observations. When the number of observations is very large, it is desirable, for the sake of computational efficiency, to perform parameter estimation online. In the maximum likelihood approach, this can be achieved using stochastic gradient search, where the corresponding gradient estimator is a non-linear functional of the optimal filter and its derivative — see, e.g., [16], [17], [20], [22]. Since the optimal filter derivative is analytically intractable for non-linear state-space models, it needs to be approximated numerically. To the best of our knowledge, only the particle approximations to the optimal filter derivative proposed in [20] and [22] are numerically stable. In the particle estimator proposed in [20], the average iteration complexity is linear in the number of particles, while the iteration running times are random. In [20], concentration inequalities and a central limit theorem have been shown for this scheme. The particle estimator proposed in [22] has quadratic iteration complexity, deterministic iteration running times and lower variance than the estimator in [20]. In [11], L_p error bounds and a central limit

*School of Mathematics, University of Bristol, Bristol, United Kingdom (v.b.tadic@bristol.ac.uk).

†Department of Statistics, University of Oxford, Oxford, United Kingdom (doucet@stats.ox.ac.uk).

theorem have been established for the scheme proposed in [22].

In this paper, we analyze the bias of the particle approximation to the optimal filter derivative proposed in [22]. Using the stability properties of the optimal filter and its derivative, we derive bounds on this bias in terms of the number of particles. These bounds cover several classes of state-space models met in practice. Moreover, under mixing conditions, these bounds are uniform in time and inversely proportional to the number of particles. To the best of our knowledge, the results presented here are the first results on the bias of the particle approximation to the optimal filter derivative proposed in [22]. They are also one of the first and most important stepping stones to analyze the asymptotic properties of online maximum likelihood estimation in non-linear state-space models — see [25].

The rest of this paper is organized as follows. In Section 2, we define the optimal filter derivative and its particle approximation. In the same section, we present the main results of the paper. These results are proved in Sections 3 – 5. Additional results are established in the Supplementary Material available at .

2. Main Results.

2.1. State-Space Models, Optimal Filter and Optimal Filter Derivative.

To define state-space models and state the optimal filtering problem, we use the following notation. For a set \mathcal{Z} in a finite dimensional space, $\mathcal{B}(\mathcal{Z})$ denotes the collection of Borel subsets of \mathcal{Z} . $d_x \geq 1$ and $d_y \geq 1$ are integers, while $\mathcal{X} \in \mathcal{B}(\mathbb{R}^{d_x})$ and $\mathcal{Y} \in \mathcal{B}(\mathbb{R}^{d_y})$. Let (Ω, \mathcal{F}, P) be a probability space. A state-space model can be described as an $\mathcal{X} \times \mathcal{Y}$ -valued stochastic process $\{(X_n, Y_n)\}_{n \geq 0}$ defined on (Ω, \mathcal{F}, P) , where the process $\{X_n\}_{n \geq 0}$ is unobservable and any information on $\{X_n\}_{n \geq 0}$ is only available through the observation process $\{Y_n\}_{n \geq 0}$. In this context, random variables X_n and Y_n are (respectively) called the state and observation at discrete-time n , while sets \mathcal{X} and \mathcal{Y} are (respectively) referred to as the state and observation spaces. One of the most important problems related to state-space models is the estimation of the states X_n and X_{n+1} given observations $Y_{0:n} := (Y_0, \dots, Y_n)$. This problem is known as filtering.

In the Bayesian approach, the estimation of states X_n and X_{n+1} given $Y_{0:n}$ is based on the optimal filtering distributions $P(X_n \in dx_n | Y_{0:n})$ and $P(X_{n+1} \in dx_{n+1} | Y_{0:n})$. In practice, the filtering distributions are usually evaluated using some approximate models. In this paper, we assume that the model $\{(X_n, Y_n)\}_{n \geq 0}$ can be accurately approximated by a parametric family of non-linear state-space models. To specify such a family, we rely on the following notation: $d \geq 1$ is an integer, while $\Theta \subseteq \mathbb{R}^d$ is an open set. $\mu(dx)$ and $\nu(dy)$ are positive measures on \mathcal{X} and \mathcal{Y} (respectively). $p_\theta(x'|x)$ and $q_\theta(y|x)$ are Borel-measurable functions which map $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ to $[0, \infty)$ and are probability densities in x' , y with respect to $\mu(dx)$, $\nu(dy)$. $\xi_\theta(dx)$ is a parameterized probability measure on \mathcal{X} , i.e., $\xi_\theta(B)$ maps $\theta \in \Theta$, $B \in \mathcal{B}(\mathcal{X})$ to $[0, 1]$ and is a probability measure in B and Borel-measurable in θ . With this notation, we can define a parametric family of state-space models as an $\mathcal{X} \times \mathcal{Y}$ -valued stochastic process $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$ which is defined on (Ω, \mathcal{F}, P) , parameterized by $\theta \in \Theta$ and satisfies

$$P((X_0^\theta, Y_0^\theta) \in B) = \int \int I_B(x, y) q_\theta(y|x) \nu(dy) \xi_\theta(dx),$$

$$P((X_{n+1}^\theta, Y_{n+1}^\theta) \in B | X_{0:n}^\theta, Y_{0:n}^\theta) = \int \int I_B(x, y) q_\theta(y|x) p_\theta(x | X_n^\theta) \nu(dy) \mu(dx)$$

almost surely for each $\theta \in \Theta$, $B \in \mathcal{B}(\mathcal{X} \times \mathcal{Y})$, $n \geq 0$.¹

Throughout the paper, we assume that $p_\theta(x'|x)$ and $q_\theta(y|x)$ are differentiable in θ for each $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. To show how the filtering distribution and its derivative are computed using the approximate model $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$, we use the following notation. If $k \geq 1$ is an integer, \mathcal{Z} is a finite dimensional space and $\zeta(dz)$ is a k -dimensional signed vector measure on \mathcal{Z} , then $\langle \zeta \rangle$ denotes the quantity $\langle \zeta \rangle = \zeta(\mathcal{Z})$. $w_\theta(x)$ a Borel-measurable function mapping $\theta \in \Theta$, $x \in \mathcal{X}$ to \mathbb{R}^d . $r_{\theta, \mathbf{y}}^n(x'|x)$ and $t_{\theta, \mathbf{y}}^n(x'|x)$ are the functions defined by

$$(2.1) \quad r_{\theta, \mathbf{y}}^n(x'|x) = p_\theta(x'|x)q_\theta(y_{n-1}|x), \quad t_{\theta, \mathbf{y}}^n(x'|x) = \nabla_\theta \log(r_{\theta, \mathbf{y}}^n(x'|x))$$

for $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $n \geq 1$ and a sequence $\mathbf{y} = \{y_n\}_{n \geq 0}$ in \mathcal{Y} . $r_{\theta, \mathbf{y}}^{m:n}(x_{m:n})$ and $t_{\theta, \mathbf{y}}^{m:n}(x_{m:n})$ are the functions defined by

$$(2.2) \quad r_{\theta, \mathbf{y}}^{m:m}(x_{m:m}) = 1, \quad r_{\theta, \mathbf{y}}^{m:n}(x_{m:n}) = \prod_{k=m+1}^n r_{\theta, \mathbf{y}}^k(x_k|x_{k-1}),$$

$$t_{\theta, \mathbf{y}}^{m:m}(x_{m:m}) = w_\theta(x_m), \quad t_{\theta, \mathbf{y}}^{m:n}(x_{m:n}) = w_\theta(x_m) + \sum_{k=m+1}^n t_{\theta, \mathbf{y}}^k(x_k|x_{k-1})$$

for $x_m, \dots, x_n \in \mathcal{X}$, $n > m \geq 0$. $\mathbb{R}_{\theta, \mathbf{y}}^n(dx_{0:n})$ and $\mathbb{T}_{\theta, \mathbf{y}}^n(dx_{0:n})$ are the measures defined by

$$\mathbb{R}_{\theta, \mathbf{y}}^n(A) = \int_A r_{\theta, \mathbf{y}}^{0:n}(x_{0:n})(\xi_\theta \times \mu^n)(dx_{0:n}),$$

$$\mathbb{T}_{\theta, \mathbf{y}}^n(A) = \int_A t_{\theta, \mathbf{y}}^{0:n}(x_{0:n})r_{\theta, \mathbf{y}}^{0:n}(x_{0:n})(\xi_\theta \times \mu^n)(dx_{0:n})$$

for $A \in \mathcal{B}(\mathcal{X}^{n+1})$, $n \geq 1$, where $\mu^n(dx_{1:n}) = \mu(dx_1) \cdots \mu(dx_n)$ and $(\xi_\theta \times \mu^n)(dx_{0:n}) = \xi_\theta(dx_0)\mu^n(dx_{1:n})$. $\mathbb{P}_{\theta, \mathbf{y}}^n(dx_{0:n})$ and $P_{\theta, \mathbf{y}}^n(dx)$ are the measures defined for $B \in \mathcal{B}(\mathcal{X})$ by

$$(2.3) \quad \mathbb{P}_{\theta, \mathbf{y}}^n(A) = \frac{\mathbb{R}_{\theta, \mathbf{y}}^n(A)}{\langle \mathbb{R}_{\theta, \mathbf{y}}^n \rangle}, \quad P_{\theta, \mathbf{y}}^n(B) = \mathbb{P}_{\theta, \mathbf{y}}^n(\mathcal{X}^n \times B).$$

$\mathbb{Q}_{\theta, \mathbf{y}}^n(dx_{0:n})$ and $Q_{\theta, \mathbf{y}}^n(dx)$ are the measures defined by

$$(2.4) \quad \mathbb{Q}_{\theta, \mathbf{y}}^n(A) = \frac{\mathbb{T}_{\theta, \mathbf{y}}^n(A)}{\langle \mathbb{R}_{\theta, \mathbf{y}}^n \rangle} - \mathbb{P}_{\theta, \mathbf{y}}^n(A) \frac{\langle \mathbb{T}_{\theta, \mathbf{y}}^n \rangle}{\langle \mathbb{R}_{\theta, \mathbf{y}}^n \rangle}, \quad Q_{\theta, \mathbf{y}}^n(B) = \mathbb{Q}_{\theta, \mathbf{y}}^n(\mathcal{X}^n \times B).$$

All results presented in this paper are based on Assumptions 2.1 – 2.3 (see Subsection 2.3, below). Using elementary calculus, it can be shown that the functions and measures defined above are well-defined under these assumptions. It can also be verified

$$(2.5) \quad P_{\theta, \mathbf{y}}^n(B) = P(X_n^\theta \in B | Y_{0:n-1}^\theta = y_{0:n-1}).$$

¹To evaluate the values of θ for which $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$ provides the best approximation to $\{(X_n, Y_n)\}_{n \geq 0}$, we usually rely on the maximum likelihood principle. For further details on maximum likelihood estimation in state-space and hidden Markov models, see, e.g., [3], [14].

Hence, $P_{\theta, \mathbf{y}}^n(dx)$ is the optimal filter (i.e., one-step predictor) for the model $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$.

Let $\lambda(dx)$ be a finite positive measure on \mathcal{X} . Moreover, let $l_\theta(x)$ be a Borel-measurable function which maps $\theta \in \Theta$, $x \in \mathcal{X}$ to $(0, \infty)$ and satisfies the following conditions: (i) $l_\theta(x)$ is differentiable in θ for each $\theta \in \Theta$, $x \in \mathcal{X}$, and (ii) $\|\nabla_\theta l_\theta(x)\|$ is uniformly bounded in (θ, x) on $\Theta \times \mathcal{X}$. Suppose that Assumptions 2.1 – 2.3 hold and that $\xi_\theta(dx)$, $w_\theta(x)$ are of the form

$$(2.6) \quad \xi_\theta(B) = \int_B l_\theta(x') \lambda(dx'), \quad w_\theta(x) = \nabla_\theta \log(l_\theta(x)).$$

Then, it is straightforward to verify

$$(2.7) \quad Q_{\theta, \mathbf{y}}^n(B) = \nabla_\theta P(X_n^\theta \in B | Y_{0:n-1}^\theta = y_{0:n-1}).$$

Thus, $Q_{\theta, \mathbf{y}}^n(B)$ is the optimal filter derivative — see [11, Section 2] for further details.

2.2. Particle Approximation to Optimal Filter Derivative. Unless the model $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$ is linear Gaussian (or the state-space \mathcal{X} has finitely many elements), $P_{\theta, \mathbf{y}}^n(dx)$ and its gradient $Q_{\theta, \mathbf{y}}^n(dx)$ do not admit closed-form expressions and need to be approximated numerically.

For a given $\theta \in \Theta$, the particle method proposed in [22] approximates $P_{\theta, \mathbf{y}}^n(dx)$ and $Q_{\theta, \mathbf{y}}^n(dx)$ respectively by the empirical distributions

$$(2.8) \quad \hat{\xi}_n^\theta(dx) = \frac{1}{N} \sum_{i=1}^N \delta_{\hat{X}_{n,i}^\theta}(dx), \quad \hat{\zeta}_n^\theta(dx) = \frac{1}{N} \sum_{i=1}^N \left(W_{n,i}^\theta - \frac{1}{N} \sum_{j=1}^N W_{n,j}^\theta \right) \delta_{\hat{X}_{n,i}^\theta}(dx).$$

Here, $N \geq 2$ is a fixed integer and $\{W_{n,i}^\theta : n \geq 0, 1 \leq i \leq N\}$ are random vectors generated through the recursion

$$(2.9) \quad W_{n+1,i}^\theta = \frac{\sum_{j=1}^N \left(p_\theta(\hat{X}_{n+1,j}^\theta | \hat{X}_{n,j}^\theta) \nabla_\theta q_\theta(Y_n | \hat{X}_{n,j}^\theta) + \nabla_\theta p_\theta(\hat{X}_{n+1,j}^\theta | \hat{X}_{n,j}^\theta) q_\theta(Y_n | \hat{X}_{n,j}^\theta) \right)}{\sum_{j=1}^N p_\theta(\hat{X}_{n+1,j}^\theta | \hat{X}_{n,j}^\theta) q_\theta(Y_n | \hat{X}_{n,j}^\theta)} + \frac{\sum_{j=1}^N p_\theta(\hat{X}_{n+1,j}^\theta | \hat{X}_{n,j}^\theta) q_\theta(Y_n | \hat{X}_{n,j}^\theta) W_{n,j}^\theta}{\sum_{j=1}^N p_\theta(\hat{X}_{n+1,j}^\theta | \hat{X}_{n,j}^\theta) q_\theta(Y_n | \hat{X}_{n,j}^\theta)},$$

where $\{\hat{X}_{n,i}^\theta : n \geq 0, 1 \leq i \leq N\}$ are random samples called particles generated through

$$(2.10) \quad \hat{X}_{n+1,i}^\theta \sim \frac{\sum_{j=1}^N p_\theta(x | \hat{X}_{n,j}^\theta) q_\theta(Y_n | \hat{X}_{n,j}^\theta) \mu(dx)}{\sum_{j=1}^N q_\theta(Y_n | \hat{X}_{n,j}^\theta)}.$$

In recursion (2.9), $\{W_{0,i}^\theta : 1 \leq i \leq N\}$ are selected as $W_{0,i}^\theta = w_\theta(\hat{X}_{0,i}^\theta)$. In recursion (2.10), $\{\hat{X}_{n+1,i}^\theta : 1 \leq i \leq N\}$ are sampled independently from one another. In the same recursion, $\{\hat{X}_{0,i}^\theta : 1 \leq i \leq N\}$ are sampled from $\xi_\theta(dx)$ independently one from another and independently from Y_0 .

REMARK. Let $w_{\theta, \mathbf{y}}^n(x)$ be the function defined by

$$w_{\theta, \mathbf{y}}^n(x) = E(t_{\theta, \mathbf{y}}^{0:n}(X_{0:n}^\theta) | X_n^\theta = x, Y_{0:n-1}^\theta = y_{0:n-1})$$

for $\theta \in \Theta$, $x \in \mathcal{X}$, $n \geq 1$ and a sequence $\mathbf{y} = \{y_n\}_{n \geq 0}$ in \mathcal{Y} . Then, we have $W_{n,i}^\theta = w_{\theta, \mathbf{Y}}^n(\hat{X}_{n,i}^\theta)$ for each $1 \leq i \leq N$, $n \geq 1$, where $\mathbf{Y} = \{Y_n\}_{n \geq 0}$. The equivalence between this representation of random vector $W_{n,i}^\theta$ and recursion (2.9) is shown and discussed in [11, Section 3]. Recursion (2.9) is derived in [22, Section 2.2].

2.3. Bias of Particle Approximation to Optimal Filter Derivative. We analyze here the bias of the particle approximations (2.8). The analysis relies on the following notation. For $z \in \mathbb{R}^k$, $k \geq 1$, $\|z\|$ denotes the l_∞ norm of z . If $\psi_\theta(x)$ is a Borel-measurable function mapping $\theta \in \Theta$, $x \in \mathcal{X}$ to \mathbb{R}^k , then $\|\psi_\theta\|$ denotes the L_∞ norm of $\psi_\theta(x)$ in x , i.e., $\|\psi_\theta\| = \sup_{x \in \mathcal{X}} \|\psi_\theta(x)\|$. If $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ is a Borel-measurable function and $\zeta(dx)$ is a k -dimensional signed vector measure on \mathcal{X} , then $\zeta(\varphi)$ denotes the integral $\zeta(\varphi) = \int \varphi(x) \zeta(dx)$.

The analysis carried out in this paper relies on the following assumptions.

ASSUMPTION 2.1. *There exists a real number $\varepsilon \in (0, 1)$ such that*

$$\varepsilon \leq p_\theta(x'|x) \leq \frac{1}{\varepsilon}, \quad \varepsilon \leq q_\theta(y|x) \leq \frac{1}{\varepsilon}$$

for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

ASSUMPTION 2.2. *There exists a real number $K \in [1, \infty)$ such that*

$$\max\{\|\nabla_\theta p_\theta(x'|x)\|, \|\nabla_\theta q_\theta(y|x)\|\} \leq K$$

for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

ASSUMPTION 2.3. $\|w_\theta\| = \sup_{x \in \mathcal{X}} \|w_\theta(x)\| < \infty$ for all $\theta \in \Theta$.

Assumption 2.1 is a standard strong mixing condition and is a crucial ingredient of many results on optimal filtering and statistical inference in state-space and hidden Markov models — see e.g., [3], [5], [11], [13], [18], [19], [21] [23]. Together with Assumption 2.2, it ensures that the optimal filter and its derivative forget initial conditions exponentially fast — see Proposition 4.1, Section 4. This assumption, together with Assumptions 2.2 and 2.3, also ensures the stability of particle approximations (2.8) — see Proposition 5.1, Section 5. Assumption 2.1 is restrictive as it implicitly requires the state and observation spaces \mathcal{X} and \mathcal{Y} to be bounded.

Let \mathbf{Y} denote stochastic process $\{Y_n\}_{n \geq 0}$, i.e., $\mathbf{Y} = \{Y_n\}_{n \geq 0}$. The main results of our paper are stated in the next theorem.

THEOREM 2.1. *Let θ be any element of Θ , while $\mathbf{y} = \{y_n\}_{n \geq 0}$ is any sequence in \mathcal{Y} . Moreover, let $\varphi : \mathcal{X} \rightarrow [-1, 1]$ be any Borel-measurable function, while n is any positive integer.*

(i) *Suppose that Assumption 2.1 holds. Then, there exists a real number $L \in [1, \infty)$ (independent of N , θ , \mathbf{y} , $\varphi(x)$, n and depending only on ε) such that*

$$(2.11) \quad \left| E \left(\hat{\xi}_n^\theta(\varphi) - P_{\theta, \mathbf{Y}}^n(\varphi) \mid \mathbf{Y} = \mathbf{y} \right) \right| \leq \frac{L}{N}.$$

(ii) *Suppose that Assumptions 2.1 – 2.3 hold. Then, there exist real numbers $\rho \in (0, 1)$, $M \in [1, \infty)$ (independent of N , θ , \mathbf{y} , $\varphi(x)$, n and depending only on ε , d , K) such that*

$$(2.12) \quad \left\| E \left(\hat{\xi}_n^\theta(\varphi) - Q_{\theta, \mathbf{Y}}^n(\varphi) \mid \mathbf{Y} = \mathbf{y} \right) \right\| \leq \frac{M(1 + \rho^n \|w_\theta\|)}{N}.$$

The proof of Theorem 2.1 is provided in Section 6 — see Proposition 6.4.

The empirical measures $\hat{\xi}_n^\theta(dx)$ and $\hat{\zeta}_n^\theta(dx)$ are estimators of the optimal predictor $P_{\theta, \mathbf{Y}}^n(dx)$ and its gradient $Q_{\theta, \mathbf{Y}}^n(dx)$. Hence, the conditional expectations in (2.11), (2.12) can be viewed as the bias of particle approximations (2.8) for which Theorem 2.1 provides bounds. These bounds are inversely proportional to N and uniform in discrete-time n as $\rho^n \leq 1$. They depend on $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$ through constants ρ , L , M and the initial conditions in recursion (2.9) (through $\|w_\theta\|$).

Due to their practical and theoretical importance, particle methods have extensively been studied in a number of papers and books — see e.g., [1], [3], [4], [6], [7], [14], [15], [14] – [16]. Within a broader analysis of the propagation of chaos in Feynman-Kac models, the bias of particle approximations to the optimal filter has been addressed in [6] – [10], [12]. Under conditions similar or identical to Assumption 2.1, the results of [6] – [10], [12] lead to Part (i) of Theorem 2.1.² As opposed to particle approximations to the optimal filter, the optimal filter derivative and its particle approximations have attracted much less attention. Part (ii) of Theorem 2.1 fills this gap in the literature on optimal filtering and particle methods. To the best of our knowledge, Part (ii) of Theorem 2.1 is the first result on the bias of the particle approximation (2.8) – (2.10). In [25], we use this result, together with the results of [24], to analyze the asymptotic behavior of recursive maximum likelihood estimation in non-linear state-space models.

3. Results Related to Empirical Measures. In this section, we present an auxiliary result on the ratio of integrals approximated using empirical measures. This result is a crucial ingredient in the proof of Lemma 6.3 which itself is a corner-stone of the main results — see Proposition 6.4. This result has already appeared in [11, Lemma A.1] and [12, Lemmas B.3, B.4]. For completeness, a proof is provided in the supplementary material (Section SM1).

We use the following additional notation. \mathcal{Z} is a finite dimensional space, while $\xi(dz)$ is a probability measure on \mathcal{Z} . $\{Z_k\}_{k \geq 1}$ are independent \mathcal{Z} -valued random variables which are defined on a probability space (Ω, \mathcal{F}, P) and distributed according to $\xi(dz)$ (i.e., $P(Z_k \in B) = \xi(B)$ for each $B \in \mathcal{B}(\mathcal{Z})$). $\xi_k(dz)$ is the empirical measure defined for $k \geq 1$ by

$$\xi_k(B) = \frac{1}{k} \sum_{i=1}^k \delta_{Z_i}(B).$$

PROPOSITION 3.1. *Let $f : \mathcal{Z} \rightarrow \mathbb{R}$ and $g : \mathcal{Z} \rightarrow (0, \infty)$ be Borel-measurable functions such that*

$$\sup_{z \in \mathcal{Z}} |f(z)| < \infty, \quad \sup_{z \in \mathcal{Z}} g(z) < \infty, \quad \inf_{z \in \mathcal{Z}} g(z) > 0.$$

Then, we have

$$(3.1) \quad \left| E \left(\frac{\xi_k(f)}{\xi_k(g)} \right) - \frac{\xi(f)}{\xi(g)} \right| \leq \frac{2\alpha\beta^2}{k}, \quad \left(E \left(\left| \frac{\xi_k(f)}{\xi_k(g)} - \frac{\xi(f)}{\xi(g)} \right|^2 \right) \right)^{1/2} \leq \frac{2\alpha\beta}{\sqrt{k}}$$

²Although Part (i) of Theorem 2.1 is a particular case in the analysis carried out in [6], [7], [8], [9], [10], [12], we include it in the main results for the following reasons: (i) $\xi_n^\theta(dx)$ is an integral part of the particle approximation (2.8) – (2.10), (ii) the bound (2.11) is an essential prerequisite for Part (ii) of Theorem 2.1, and (iii) the proof of Part (i) of Theorem 2.1 presented here seems more direct than the analysis carried out in [6], [7], [8], [9], [10], [12].

for any $k \geq 1$, where α, β are defined by

$$(3.2) \quad \alpha = \sup_{z', z'' \in \mathcal{Z}} \left| \frac{f(z')}{g(z')} - \frac{f(z'')}{g(z'')} \right|, \quad \beta = \sup_{z', z'' \in \mathcal{Z}} \frac{g(z')}{g(z'')}.$$

4. Results Related to Stability of Optimal Filter and Its Derivative.

In this section, we present results on the stability properties of the optimal predictor $P_{\theta, \mathbf{y}}^n(dx)$ and its gradient $Q_{\theta, \mathbf{y}}^{m:n}(dx)$. These results are prerequisites for the proof of the main results — see Lemmas 6.2, 6.3 and Proposition 6.4.

The following additional notation is used here. $\mathcal{P}(\mathcal{X})$ is the collection of probability measures on \mathcal{X} , while the set of Borel-measurable functions mapping \mathcal{X} to \mathbb{R} is denoted by $\mathcal{F}(\mathcal{X})$. $\mathcal{M}_p(\mathcal{X})$ is the set of positive measures on \mathcal{X} , while the collection of signed measures on \mathcal{X} is denoted by $\mathcal{M}_s(\mathcal{X})$. If $k \geq 1$ is an integer, then $\mathcal{M}_s^k(\mathcal{X})$ is the set of k -dimensional signed vector measures on \mathcal{X} and $\mathcal{F}^k(\mathcal{X})$ is the collection of Borel-measurable functions mapping \mathcal{X} to \mathbb{R}^k . For $\xi \in \mathcal{M}_s(\mathcal{X})$, $|\xi|(dx)$ and $\|\xi\|$ denote (respectively) the total variation and the total variation norm of $\xi(dx)$. For $\zeta \in \mathcal{M}_s^k(\mathcal{X})$, $|\zeta|(dx)$ and $\|\zeta\|$ denote (respectively) the total variation and the total variation norm of $\zeta(dx)$ induced by l_1 vector norm.³ If $H(\zeta)$ is a function mapping $\zeta \in \mathcal{M}_s^k$ to \mathcal{M}_s^l , then $H(\zeta)(B)$ stands for the measure of $B \in \mathcal{B}(\mathcal{X})$ with respect to $H(\zeta)$. Moreover, if $\xi \in \mathcal{M}_s(\mathcal{X})$, $\zeta \in \mathcal{M}_s^k(\mathcal{X})$ and $R(x, dx')$, $S(x, dx')$ are integral operators from $\mathcal{F}(\mathcal{X})$ to $\mathcal{F}(\mathcal{X})$, $\mathcal{F}^k(\mathcal{X})$ (respectively), then $(\zeta R)(dx)$ and $(\xi S)(dx)$ denote the elements of $\mathcal{M}_s^k(\mathcal{X})$ defined for $B \in \mathcal{B}(\mathcal{X})$ by

$$(\zeta R)(B) = \int_B R(x, B) \zeta(dx), \quad (\xi S)(B) = \int_B S(x, B) \xi(dx).$$

$s_{\theta, \mathbf{y}}^{m:n}(x_{m:n})$ is the function defined by

$$(4.1) \quad s_{\theta, \mathbf{y}}^{m:m}(x_{m:m}) = 0, \quad s_{\theta, \mathbf{y}}^{m:n}(x_{m:n}) = \nabla_{\theta} r_{\theta, \mathbf{y}}^{m:n}(x_{m:n})$$

for $\theta \in \Theta$, $x_m, \dots, x_n \in \mathcal{X}$, $n > m \geq 0$ and a sequence $\mathbf{y} = \{y_n\}_{n \geq 0}$ in \mathcal{Y} ($r_{\theta, \mathbf{y}}^{m:n}(x_{m:n})$ is specified in (2.2)). $R_{\theta, \mathbf{y}}^{m:n}(x, dx')$ and $S_{\theta, \mathbf{y}}^{m:n}(x, dx')$ are the integral operators from $\mathcal{F}(\mathcal{X})$ to $\mathcal{F}(\mathcal{X})$, $\mathcal{F}^d(\mathcal{X})$ (respectively) defined by

$$(4.2) \quad R_{\theta, \mathbf{y}}^{m:m}(x, B) = \delta_x(B), \quad R_{\theta, \mathbf{y}}^{m:n}(x, B) = \int_{\mathcal{X}^{n-m} \times B} r_{\theta, \mathbf{y}}^{m:n}(x_{m:n}) (\delta_x \times \mu^{n-m})(dx_{m:n}),$$

$$(4.3) \quad S_{\theta, \mathbf{y}}^{m:m}(x, B) = 0, \quad S_{\theta, \mathbf{y}}^{m:n}(x, B) = \int_{\mathcal{X}^{n-m} \times B} s_{\theta, \mathbf{y}}^{m:n}(x_{m:n}) (\delta_x \times \mu^{n-m})(dx_{m:n}),$$

where $(\delta_x \times \mu^{n-m})(dx_{m:n}) = \delta_x(dx_m) \mu(dx_{m+1}) \cdots \mu(dx_n)$. $F_{\theta, \mathbf{y}}^{m:n}(\xi)$, $G_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)$ and $H_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)$ are the functions mapping $\xi \in \mathcal{P}(\mathcal{X})$, $\zeta \in \mathcal{M}_s^d(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$, $\mathcal{M}_s^d(\mathcal{X})$,

³If $\zeta \in \mathcal{M}_s^k(\mathcal{X})$, then $|\zeta|(dx) = \sum_{i=1}^k |e_i^T \zeta|(dx)$ and $\|\zeta\| = \sum_{i=1}^k \|e_i^T \zeta\|$, where e_i is the i -th standard unit vector in \mathbb{R}^k .

$\mathcal{M}_s^d(\mathcal{X})$ (respectively) defined by

(4.4)

$$F_{\theta, \mathbf{y}}^{m:m}(\xi)(B) = \xi(B), \quad F_{\theta, \mathbf{y}}^{m:n}(\xi)(B) = \frac{(\xi R_{\theta, \mathbf{y}}^{m:n})(B)}{\langle \xi R_{\theta, \mathbf{y}}^{m:n} \rangle},$$

(4.5)

$$H_{\theta, \mathbf{y}}^{m:m}(\xi, \zeta)(B) = \zeta(B), \quad H_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)(B) = \frac{(\zeta R_{\theta, \mathbf{y}}^{m:n})(B) + (\xi S_{\theta, \mathbf{y}}^{m:n})(B)}{\langle \xi R_{\theta, \mathbf{y}}^{m:n} \rangle},$$

(4.6)

$$G_{\theta, \mathbf{y}}^{m:m}(\xi, \zeta)(B) = \zeta(B), \quad G_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)(B) = H_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)(B) - F_{\theta, \mathbf{y}}^{m:n}(\xi)(B) \langle H_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta) \rangle.$$

REMARK. For $\theta \in \Theta$, let $\zeta_\theta(dx)$ be the element of $\mathcal{M}_s^d(\mathcal{X})$ specified in (6.1) (below). Suppose that Assumptions 2.1 – 2.3 hold. Then, it easy to show

(4.7)

$$P_{\theta, \mathbf{y}}^n(B) = F_{\theta, \mathbf{y}}^{0:n}(\xi_\theta)(B), \quad Q_{\theta, \mathbf{y}}^n(B) = G_{\theta, \mathbf{y}}^{0:n}(\xi_\theta, \zeta_\theta)(B)$$

for $n \geq 1$. Given (2.5), (2.7), (4.7), $F_{\theta, \mathbf{y}}^{m:n}(\xi)$ and $G_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)$ can be considered as a generalization of the optimal filter and its gradient. In this context, ξ, ζ can be viewed as initial conditions in the recursion generating the optimal filter and its gradient (for further details, see [18], [23]).

PROPOSITION 4.1. Let θ be any element of Θ , while $\mathbf{y} = \{y_n\}_{n \geq 0}$ is any sequence in \mathcal{Y} . Moreover, let ξ, ξ' be any elements of $\mathcal{P}(\mathcal{X})$, while ζ, ζ' are any elements of $\mathcal{M}_s^d(\mathcal{X})$. Further to this, let n, m be any integers satisfying $n \geq m \geq 0$.

(i) Suppose that Assumption 2.1 holds. Then, there exist real numbers $\rho_1 \in (0, 1)$, $C_1 \in [1, \infty)$ (independent of $\theta, \mathbf{y}, \xi, \xi', n, m$ and depending only on ε) such that

$$(4.8) \quad \|F_{\theta, \mathbf{y}}^{m:n}(\xi) - F_{\theta, \mathbf{y}}^{m:n}(\xi')\| \leq C_1 \rho_1^{n-m}, \quad \frac{\langle \xi R_{\theta, \mathbf{y}}^{m:n} \rangle}{\langle \xi' R_{\theta, \mathbf{y}}^{m:n} \rangle} \leq C_1.$$

(ii) Suppose that Assumptions 2.1 and 2.2 hold. Then, there exist real numbers $\rho_2 \in (0, 1)$, $C_2 \in [1, \infty)$ (independent of $\theta, \mathbf{y}, \xi, \xi', \zeta, \zeta', n, m$ and depending only on ε, d, K) such that

$$(4.9) \quad \|G_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta) - G_{\theta, \mathbf{y}}^{m:n}(\xi', \zeta')\| \leq C_2 \rho_2^{n-m} (1 + \|\zeta\| + \|\zeta'\|),$$

$$(4.10) \quad \|H_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)\| \leq C_2 (n - m + \|\zeta\|).$$

Proposition 4.1 is a relatively straightforward extension of the results of [18], [19], [23] to the optimal predictor and its gradient. A detailed proof of the proposition is provided in the supplementary material (Section SM2).

5. Results Related to Stability of Particle Approximations. In this section, we consider the particle approximation $\hat{\zeta}_n^\theta(dx)$ and its stability. Using results on the (Dobrushin) ergodicity coefficient, we show that the sequence $\{\|\hat{\zeta}_n^\theta\|\}_{n \geq 0}$ is bounded uniformly in θ . The results presented here are prerequisites for the proof of the main results — see Lemma 6.3 and Proposition 6.4.

To state the results of this section, additional notation needs to be introduced. \mathcal{P}^N is the set of N -dimensional probability vectors. $\mathcal{P}^{N \times N}$ is the set of $N \times N$ (column) stochastic matrices (i.e., $A \in \mathcal{P}^{N \times N}$ if and only if the columns of A are elements of \mathcal{P}^N). e is the element of \mathbb{R}^N whose all elements are one. For $1 \leq i \leq N$,

e_i is the i -th standard unit vector in \mathbb{R}^N (i.e., e_i is the element of \mathcal{P}^N whose i -th element is one). For $z \in \mathbb{R}^N$, $\|z\|_1$ and $\|z\|$ are (respectively) the l_1 and l_∞ norm of z . For $B \in \mathbb{R}^{d \times N}$, $\|B\|$ is the l_∞ norm of B (i.e., $\|B\|$ is the maximum absolute value of the entries of B). For $A \in \mathcal{P}^{N \times N}$, $\tau(A)$ is the ergodicity coefficient of A , i.e.,

$$(5.1) \quad \tau(A) = \frac{1}{2} \max_{1 \leq j', j'' \leq N} \sum_{i=1}^N |A_{i,j'} - A_{i,j''}| = 1 - \min_{1 \leq j', j'' \leq N} \sum_{i=1}^N \min\{A_{i,j'}, A_{i,j''}\},$$

where $A_{i,j}$ is the (i, j) entry of A (see [2, Section 15.2.1] for more details on the ergodicity coefficient and its equivalent forms). A_n^θ and B_n^θ are (respectively) the $N \times N$ and $d \times N$ random matrices defined by

$$(5.2) \quad A_{n,i,j}^\theta = \frac{r_{\theta, \mathbf{Y}}^n(\hat{X}_{n,j}^\theta | \hat{X}_{n-1,i}^\theta)}{\sum_{k=1}^N r_{\theta, \mathbf{Y}}^n(\hat{X}_{n,j}^\theta | \hat{X}_{n-1,k}^\theta)}, \quad B_{n,j}^\theta = \frac{\sum_{k=1}^N \nabla_{\theta} r_{\theta, \mathbf{Y}}^n(\hat{X}_{n,j}^\theta | \hat{X}_{n-1,k}^\theta)}{\sum_{k=1}^N r_{\theta, \mathbf{Y}}^n(\hat{X}_{n,j}^\theta | \hat{X}_{n-1,k}^\theta)},$$

where $A_{n,i,j}^\theta$ is the (i, j) entry of A_n^θ and $B_{n,j}^\theta$ is the j -th column of B_n^θ . $r_{\theta, \mathbf{Y}}^n(x' | x)$ and \mathbf{Y} are specified in (2.1) and Subsection 2.3 (respectively). V_n^θ , W_n^θ and $V_{n,i}^\theta$ are the $d \times N$ random matrices and the d -dimensional random vector defined by

$$(5.3) \quad V_{n,i}^\theta = W_{n,i}^\theta - \frac{1}{N} \sum_{j=1}^N W_{n,j}^\theta, \quad V_n^\theta = (V_{n,1}^\theta, \dots, V_{n,N}^\theta), \quad W_n^\theta = (W_{n,1}^\theta, \dots, W_{n,N}^\theta)$$

for $n \geq 0$. Notice here that $V_{n,i}^\theta$ and $W_{n,i}^\theta$ are the i -th columns of V_n^θ and W_n^θ . Then, it is easy to show $A_n^\theta \in \mathcal{P}^{N \times N}$ and

$$(5.4) \quad V_n^\theta = W_n^\theta \left(I - \frac{ee^T}{N} \right), \quad W_{n+1}^\theta = W_n^\theta A_{n+1}^\theta + B_{n+1}^\theta,$$

where I is the $N \times N$ unit matrix.

REMARK. Throughout this and subsequent sections, the following convention is applied. Diacritic \sim is used to denote a locally defined quantity, i.e., a quantity whose definition holds only within the proof where the quantity appears.

PROPOSITION 5.1. Let θ be any element of Θ , while n is any non-negative integer. Suppose that Assumptions 2.1 – 2.3 hold. Then, there exist real numbers $\rho_3 \in (0, 1)$, $C_3 \in [1, \infty)$ (independent of N , θ , n and depending only on ε , d , K) such that

$$(5.5) \quad \|\hat{\zeta}_n^\theta\| \leq C_3 (1 + \rho_3^n \|w_\theta\|).$$

Proof. Throughout the proof, the following notation is used. ρ_3 , C_3 are the real numbers defined by $\rho_3 = 1 - \varepsilon^4$, $C_3 = 8Kd\varepsilon^{-7}$ (ε , K are specified in Assumptions 2.1 and 2.2). $\tilde{A}_{k,l}^\theta$ is the matrix defined by

$$(5.6) \quad \tilde{A}_{k,k}^\theta = I, \quad \tilde{A}_{k,l}^\theta = A_{k+1}^\theta \cdots A_l^\theta$$

for $l > k \geq 0$.

Iterating the second part of (5.4), we get

$$(5.7) \quad W_n^\theta = W_0^\theta \tilde{A}_{0,n}^\theta + \sum_{k=1}^n B_k^\theta \tilde{A}_{k,n}^\theta$$

for $n \geq 1$. Since $\tilde{A}_{0,n}^\theta \in \mathcal{P}^{N \times N}$, we also have $e^T \tilde{A}_{0,n}^\theta = e^T$. Consequently, the first part of (5.4) implies

$$\begin{aligned} V_0^\theta \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right) &= W_0^\theta \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right) - \frac{W_0^\theta e}{N} e^T \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right) \\ &= W_0^\theta \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right). \end{aligned}$$

Combining this with the first part of (5.4) and (5.7), we get

$$\begin{aligned} (5.8) \quad V_n^\theta &= W_0^\theta \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right) + \sum_{k=1}^n B_k^\theta \tilde{A}_{k,n}^\theta \left(I - \frac{ee^T}{N} \right) \\ &= V_0^\theta \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right) + \sum_{k=1}^n B_k^\theta \tilde{A}_{k,n}^\theta \left(I - \frac{ee^T}{N} \right). \end{aligned}$$

Owing to Assumptions 2.1, 2.2, we have $\varepsilon^2 \leq r_{\theta, \mathbf{Y}}^k(\hat{X}_{k,j}^\theta | \hat{X}_{k-1,i}^\theta) \leq 1/\varepsilon^2$ and

$$\begin{aligned} (5.9) \quad \|\nabla_{\theta} r_{\theta, \mathbf{Y}}^k(\hat{X}_{k,j}^\theta | \hat{X}_{k-1,i}^\theta)\| &\leq q_{\theta}(Y_{k-1} | \hat{X}_{k-1,i}^\theta) \|\nabla_{\theta} p_{\theta}(\hat{X}_{k,j}^\theta | \hat{X}_{k-1,i}^\theta)\| \\ &\quad + p_{\theta}(\hat{X}_{k,j}^\theta | \hat{X}_{k-1,i}^\theta) \|\nabla_{\theta} q_{\theta}(Y_{k-1} | \hat{X}_{k-1,i}^\theta)\| \leq \frac{2K}{\varepsilon} \end{aligned}$$

for $1 \leq i, j \leq N$, $k \geq 1$. Therefore, we get

$$N\varepsilon^2 \leq \sum_{i=1}^N r_{\theta, \mathbf{Y}}^k(\hat{X}_{k,j}^\theta | \hat{X}_{k-1,i}^\theta) \leq \frac{N}{\varepsilon^2}, \quad \sum_{i=1}^N \|\nabla_{\theta} r_{\theta, \mathbf{Y}}^k(\hat{X}_{k,j}^\theta | \hat{X}_{k-1,i}^\theta)\| \leq \frac{2KN}{\varepsilon}.$$

Consequently, (5.2) implies

$$(5.10) \quad A_{k,i,j}^\theta \geq \frac{\varepsilon^4}{N}, \quad \|B_{k,j}^\theta\| \leq \frac{2K}{\varepsilon^3}, \quad \|B_k^\theta\| = \max_{1 \leq j \leq N} \|B_{k,j}^\theta\| \leq \frac{2K}{\varepsilon^3}.$$

Hence, (5.1) yields $\tau(A_k^\theta) \leq 1 - \varepsilon^4 = \rho_3$.

Due to the well-known results in Markov chain theory (e.g. [2, Theorems 15.2.4, 15.2.5]), we have

$$(5.11) \quad \tau(A'A'') \leq \tau(A')\tau(A''), \quad \|A(z' - z'')\|_1 \leq \tau(A)\|z' - z''\|_1$$

for any $A, A', A'' \in \mathcal{P}^{N \times N}$, $z', z'' \in \mathcal{P}^N$. Then, using (1), we get $\tau(\tilde{A}_{k,k}^\theta) = 1$ and

$$\tau(\tilde{A}_{k,l}^\theta) \leq \tau(A_{k+1}^\theta) \cdots \tau(A_l^\theta) \leq \rho_3^{l-k}$$

for $l > k \geq 0$. Since $e_i, \frac{e}{N} \in \mathcal{P}^N$, we deduce

$$(5.12) \quad \left\| \tilde{A}_{k,l}^\theta \left(e_i - \frac{e}{N} \right) \right\|_1 \leq \tau(\tilde{A}_{k,l}^\theta) \left\| e_i - \frac{e}{N} \right\|_1 \leq 2\rho_3^{l-k}$$

for $1 \leq i \leq N$, $l \geq k \geq 0$. Consequently, (5.10) yields

$$\begin{aligned} \left\| V_0^\theta \tilde{A}_{0,l}^\theta \left(e_i - \frac{e}{N} \right) \right\| &\leq \|V_0^\theta\| \left\| \tilde{A}_{0,l}^\theta \left(e_i - \frac{e}{N} \right) \right\|_1 \leq 2\rho_3^l \|V_0^\theta\|, \\ \left\| B_k^\theta \tilde{A}_{k,l}^\theta \left(e_i - \frac{e}{N} \right) \right\| &\leq \|B_k^\theta\| \left\| \tilde{A}_{k,l}^\theta \left(e_i - \frac{e}{N} \right) \right\|_1 \leq \frac{4K\rho_3^{l-k}}{\varepsilon^3}. \end{aligned}$$

As vectors $B_k^\theta \tilde{A}_{k,l}^\theta (e_i - \frac{e}{N})$, $V_0^\theta \tilde{A}_{0,l}^\theta (e_i - \frac{e}{N})$ are the i -th columns of matrices $B_k^\theta \tilde{A}_{k,l}^\theta (I - \frac{ee^T}{N})$, $V_0^\theta \tilde{A}_{0,l}^\theta (I - \frac{ee^T}{N})$ (respectively), we conclude

$$\begin{aligned} \left\| V_0^\theta \tilde{A}_{0,l}^\theta \left(I - \frac{ee^T}{N} \right) \right\| &= \max_{1 \leq i \leq N} \left\| V_0^\theta \tilde{A}_{0,l}^\theta \left(e_i - \frac{e}{N} \right) \right\| \leq 2\rho_3^l \|V_0^\theta\|, \\ \left\| B_k^\theta \tilde{A}_{k,l}^\theta \left(I - \frac{ee^T}{N} \right) \right\| &= \max_{1 \leq i \leq N} \left\| B_k^\theta \tilde{A}_{k,l}^\theta \left(e_i - \frac{e}{N} \right) \right\| \leq \frac{4K\rho_3^{l-k}}{\varepsilon^3}. \end{aligned}$$

Hence, (5.8) implies

$$\begin{aligned} (5.13) \quad \|V_n^\theta\| &\leq \left\| V_0^\theta \tilde{A}_{0,n}^\theta \left(I - \frac{ee^T}{N} \right) \right\| + \sum_{k=1}^n \left\| B_k^\theta \tilde{A}_{k,n}^\theta \left(I - \frac{ee^T}{N} \right) \right\| \\ &\leq 2\rho_3^n \|V_0^\theta\| + \frac{4K}{\varepsilon^3} \sum_{k=1}^n \rho_3^{n-k} \leq \frac{4K}{\varepsilon^7} (1 + \rho_3^n \|V_0^\theta\|) \end{aligned}$$

for $n \geq 1$. Since $W_{0,i}^\theta = w_\theta(\hat{X}_{0,i}^\theta)$, Assumption 2.3 and (5.3) yield

$$\|V_{0,i}^\theta\| \leq \|W_{0,i}^\theta\| + \frac{1}{N} \sum_{j=1}^N \|W_{0,j}^\theta\| \leq 2\|w_\theta\|$$

for $1 \leq i \leq N$. Thus, we have $\|V_0^\theta\| \leq \|w_\theta\|$. Consequently, (5.13) implies

$$\|V_{n,i}^\theta\| \leq \|V_n^\theta\| \leq \frac{8K}{\varepsilon^7} (1 + \rho_3^n \|w_\theta\|)$$

for $n \geq 0$. As $\hat{\zeta}_n^\theta(dx) = \frac{1}{N} \sum_{j=1}^N V_{n,j}^\theta \delta_{\hat{X}_{n,j}^\theta}(dx)$ (due to (2.8), (5.3)), we get

$$\|\hat{\zeta}_n^\theta(B)\| \leq \frac{1}{N} \sum_{i=1}^N \|V_{n,i}^\theta\| \leq \frac{8K}{\varepsilon^7} (1 + \rho_3^n \|w_\theta\|)$$

for $B \in \mathcal{B}(\mathcal{X})$. Hence, we have

$$\|\hat{\zeta}_n\| \leq \frac{8dK}{\varepsilon^7} (1 + \rho_3^n \|w_\theta\|) = C_3 (1 + \rho_3^n \|w_\theta\|). \quad \square$$

6. Proof of Main Results. In this section, Proposition 6.4 is proved, while Theorem 2.1 directly follows from it. Lemma 6.3 and decompositions (6.15), (6.67), (6.72) can be considered as the corner-stones in the proof of Proposition 6.4 — see inequalities (6.68) – (6.71), (6.73), (6.74). Proposition 3.1, conditional distributions (6.41), (6.42) and identities (6.43), (6.48), (6.49) are the main ingredients in the proof of Lemma 6.3 — see inequalities (6.46), (6.47), (6.51), (6.52), (6.57). Propositions 4.1, 5.1 and Lemma 6.1 are important ingredients of the proof of Lemma 6.3, too — see inequalities (6.44), (6.45), (6.50), (6.53), (6.55). Proposition 4.1 plays an important role in the proof of Lemma 6.3, either — see inequalities (6.26) – (6.31).

Throughout this section, the following notation is used: $u_\theta(x)$, \bar{w}_θ and $\zeta_\theta(dx)$ are the functions and the element of $\mathcal{M}_s^d(\mathcal{X})$ defined by

$$(6.1) \quad \bar{w}_\theta = \int w_\theta(x') \xi_\theta(dx'), \quad u_\theta(x) = w_\theta(x) - \bar{w}_\theta, \quad \zeta_\theta(B) = \int_B u_\theta(x') \xi_\theta(dx')$$

for $\theta \in \Theta$, $x \in \mathcal{X}$, $B \in \mathcal{B}(\mathcal{X})$. $\hat{\xi}_{-1}^\theta(dx)$ and $\hat{\zeta}_{-1}^\theta(dx)$ are the elements of $\mathcal{P}(\mathcal{X})$ and $\mathcal{M}_s^d(\mathcal{X})$ (respectively) defined by

$$(6.2) \quad \hat{\xi}_{-1}^\theta(B) = \xi_\theta(B), \quad \hat{\zeta}_{-1}^\theta(B) = \zeta_\theta(B).$$

$\hat{v}_n^\theta(x)$ is the (random) function defined by

$$(6.3) \quad \hat{v}_0^\theta(x) = u_\theta(x), \quad \hat{v}_n^\theta(x) = \frac{\int r_{\theta, \mathbf{Y}}^n(x|x') \hat{\zeta}_{n-1}^\theta(dx') + \int \nabla_\theta r_{\theta, \mathbf{Y}}^n(x|x') \hat{\xi}_{n-1}^\theta(dx')}{\int r_{\theta, \mathbf{Y}}^n(x|x') \hat{\xi}_{n-1}^\theta(dx')}$$

for $n \geq 1$ ($r_{\theta, \mathbf{Y}}^n(x|x')$ and \mathbf{Y} are defined in (2.1) and Subsection 2.3, respectively). $\hat{\alpha}_n^\theta(\xi)$ is the (random) function mapping $\xi \in \mathcal{P}(\mathcal{X})$ to $\mathcal{M}_s^d(\mathcal{X})$ defined by

$$(6.4) \quad \hat{\alpha}_n^\theta(\xi)(B) = \int_B \hat{v}_n^\theta(x) \xi(dx)$$

for $\xi \in \mathcal{P}(\mathcal{X})$, $n \geq 0$. $\hat{F}_{m:n}^\theta(dx)$, $\hat{G}_{m:n}^\theta(dx)$ and $\hat{H}_{m:n}^\theta(dx)$ are the (random) elements of $\mathcal{P}(\mathcal{X})$, $\mathcal{M}_s^d(\mathcal{X})$ and $\mathcal{M}_s^d(\mathcal{X})$ (respectively) defined by

$$(6.5) \quad \hat{F}_{-1:n}^\theta(B) = F_{\theta, \mathbf{Y}}^{0:n}(\hat{\xi}_{-1}^\theta)(B), \quad \hat{F}_{m:n}^\theta(B) = F_{\theta, \mathbf{Y}}^{m:n}(\hat{\xi}_m^\theta)(B),$$

$$(6.6) \quad \hat{G}_{-1:n}^\theta(B) = G_{\theta, \mathbf{Y}}^{0:n}(\hat{\xi}_{-1}^\theta, \hat{\zeta}_{-1}^\theta)(B), \quad \hat{G}_{m:n}^\theta(B) = G_{\theta, \mathbf{Y}}^{m:n}(\hat{\xi}_m^\theta, \hat{\zeta}_m^\theta)(B),$$

$$(6.7) \quad \hat{H}_{-1:n}^\theta(B) = H_{\theta, \mathbf{Y}}^{0:n}(\hat{\xi}_{-1}^\theta, \hat{\zeta}_{-1}^\theta)(B), \quad \hat{H}_{m:n}^\theta(B) = H_{\theta, \mathbf{Y}}^{m:n}(\hat{\xi}_m^\theta, \hat{\zeta}_m^\theta)(B)$$

for $n \geq m \geq 0$ ($F_{\theta, \mathbf{y}}^{m:n}(\xi)$, $G_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)$, $H_{\theta, \mathbf{y}}^{m:n}(\xi, \zeta)$ are specified in (4.4) – (4.6)). $\hat{R}_{m:n}^\theta(x, dx')$ and $\hat{S}_{m:n}^\theta(x, dx')$ are the (random) integral operators from $\mathcal{F}(\mathcal{X})$ to $\mathcal{F}(\mathcal{X})$, $\mathcal{F}^d(\mathcal{X})$ (respectively) defined by

$$(6.8) \quad \hat{R}_{m:n}^\theta(x, B) = R_{\theta, \mathbf{Y}}^{m:n}(x, B), \quad \hat{S}_{m:n}^\theta(x, B) = S_{\theta, \mathbf{Y}}^{m:n}(x, B).$$

$\hat{\Psi}_{\theta, \mathbf{Y}}^{m:n}(x, dx')$ and $\hat{\Phi}_{\theta, \mathbf{Y}}^{m:n}(x, dx')$ are the (random) integral operators from $\mathcal{F}(\mathcal{X})$ to $\mathcal{F}^d(\mathcal{X})$ defined by

$$(6.9) \quad \hat{\Psi}_{m:n}^\theta(x, B) = \hat{R}_{m:n}^\theta(x, B) \hat{v}_m^\theta(x) + \hat{S}_{m:n}^\theta(x, B),$$

$$(6.10) \quad \hat{\Phi}_{m:n}^\theta(x, B) = \hat{\Psi}_{m:n}^\theta(x, B) - \hat{F}_{m-1:n}^\theta(B) \hat{\Psi}_{m:n}^\theta(x, \mathcal{X}).$$

$\hat{C}_{m:n}^\theta(dx)$, $\hat{B}_{m:n}^\theta(dx)$ and $\hat{A}_{m:n}^\theta(dx)$ are the (random) elements of $\mathcal{M}_s^d(\mathcal{X})$ defined by

$$(6.11) \quad \hat{C}_{m:n}^\theta(B) = \frac{(\hat{\xi}_m^\theta \hat{\Psi}_{m:n}^\theta)(B)}{\langle \hat{\xi}_m^\theta \hat{R}_{m:n}^\theta \rangle},$$

$$(6.12) \quad \hat{B}_{m:n}^\theta(B) = -(\hat{F}_{m:n}^\theta(B) - \hat{F}_{m-1:n}^\theta(B)) \langle \hat{C}_{m:n}^\theta \rangle,$$

$$(6.13) \quad \hat{A}_{m:n}^\theta(B) = \hat{C}_{m:n}^\theta(B) - \hat{F}_{m-1:n}^\theta(B) \langle \hat{C}_{m:n}^\theta \rangle.$$

LEMMA 6.1. *Let θ , B , ξ be any elements of Θ , $\mathcal{B}(\mathcal{X})$, $\mathcal{P}(\mathcal{X})$ (respectively). Moreover, let n , m be any integers satisfying $n \geq m \geq 0$.*

(i) *Suppose that Assumption 2.1 holds. Then, we have*

$$(6.14) \quad \hat{F}_{m-1:n}^\theta(B) = \frac{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(B)}{\langle \hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta \rangle}.$$

(ii) Suppose that Assumptions 2.1 – 2.3 hold. Then, we have

$$(6.15) \quad \hat{G}_{m:n}^\theta(B) = \hat{A}_{m:n}^\theta(B) + \hat{B}_{m:n}^\theta(B),$$

$$(6.16) \quad \frac{(\xi \hat{\Psi}_{m:n}^\theta)(B)}{\langle \xi \hat{R}_{m:n}^\theta \rangle} = H_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi))(B),$$

$$(6.17) \quad \frac{(\xi \hat{\Phi}_{m:n}^\theta)(B)}{\langle \xi \hat{R}_{m:n}^\theta \rangle} = (F_{\theta, \mathbf{Y}}^{m:n}(\xi)(B) - \hat{F}_{m-1:n}^\theta(B)) \langle H_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi)) \rangle \\ + G_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi))(B).$$

We also have

$$(6.18) \quad \hat{H}_{m-1:n}^\theta(B) = \frac{(\hat{F}_{m-1:m}^\theta \hat{\Psi}_{m:n}^\theta)(B)}{\langle \hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta \rangle}, \quad \hat{G}_{m-1:n}^\theta(B) = \frac{(\hat{F}_{m-1:m}^\theta \hat{\Phi}_{m:n}^\theta)(B)}{\langle \hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta \rangle}.$$

Lemma 6.1 summarizes relatively straightforward relationships between the measures defined in (6.5) – (6.7) and (6.11) – (6.13). A detailed proof of the lemma is provided in the supplementary material (Section SM3).

LEMMA 6.2. Let θ be any element of Θ , while ξ, ξ' are any elements of $\mathcal{P}(\mathcal{X})$. Moreover, let n, m be any integers satisfying $n \geq m \geq 0$. Suppose that Assumptions 2.1 and 2.2 hold. Then, there exist real numbers $\rho_4 \in (0, 1)$, $C_4 \in [1, \infty)$ (independent of θ, ξ, ξ', n, m and depending only on ε, d, K) such that

$$(6.19) \quad \max \left\{ \|\hat{C}_{m:n}^\theta\|, \|\hat{H}_{m-1:n}^\theta\| \right\} \leq C_4 (1 + n - m + \rho_4^m \|w_\theta\|),$$

$$(6.20) \quad \left\| \frac{\xi \hat{\Psi}_{m:n}^\theta}{\langle \xi \hat{R}_{m:n}^\theta \rangle} - \frac{\xi' \hat{\Psi}_{m:n}^\theta}{\langle \xi' \hat{R}_{m:n}^\theta \rangle} \right\| \leq C_4 (1 + n - m + \rho_4^m \|w_\theta\|),$$

$$(6.21) \quad \left\| \frac{\xi \hat{\Phi}_{m:n}^\theta}{\langle \xi \hat{R}_{m:n}^\theta \rangle} - \frac{\xi' \hat{\Phi}_{m:n}^\theta}{\langle \xi' \hat{R}_{m:n}^\theta \rangle} \right\| \leq C_4 \rho_4^{n-m} (1 + n - m + \rho_4^m \|w_\theta\|).$$

Proof. Throughout the proof, the following notation is used. x, x' are any elements of \mathcal{X} , while B is any element of $\mathcal{B}(\mathcal{X})$. ρ_4 is the real number defined by $\rho_4 = \max \{\rho_1, \rho_2, \rho_3\}$, while $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3, \tilde{C}_4, C_4$ are the real numbers defined as $\tilde{C}_1 = 2dC_3$, $\tilde{C}_2 = 4\tilde{C}_1 K \varepsilon^{-4} \rho_4^{-1}$, $\tilde{C}_3 = d\tilde{C}_2$, $\tilde{C}_4 = 3C_2 \tilde{C}_1 \tilde{C}_3$, $C_4 = 3C_1 \tilde{C}_4$ ($\varepsilon, \rho_1, \rho_2, K, C_1, C_2$ are specified in Assumptions 2.1, 2.2 and Proposition 4.1). n, m are any integers satisfying $n \geq m \geq 0$.

Relying on Assumption 2.3 and (6.1), (6.3), we conclude

$$(6.22) \quad \|\hat{v}_0^\theta(x)\| = \|u_\theta(x)\| \leq \|w_\theta(x)\| + \int \|w_\theta(x')\| \xi_\theta(dx') \leq 2\|w_\theta\|.$$

Consequently, (6.1), (6.2) imply

$$\|\hat{\zeta}_{-1}^\theta(B)\| = \|\zeta_\theta(B)\| \leq \int_B \|u_\theta(x)\| \xi_\theta(dx) \leq 2\|w_\theta\|.$$

Hence, we have $\|\hat{\zeta}_{-1}^\theta\| = \|\zeta_\theta\| \leq 2d\|w_\theta\|$. Combining this with Proposition 5.1, we get

$$(6.23) \quad \|\hat{\zeta}_k^\theta\| \leq 2dC_3(1 + \rho_3^k \|w_\theta\|) \leq \tilde{C}_1(1 + \rho_4^k \|w_\theta\|)$$

for $k \geq -1$.

Using Assumptions 2.1, 2.2 and the same arguments as in Proposition 5.1 (see (5.9)), we deduce

$$\varepsilon^2 \leq r_{\theta, \mathbf{Y}}^k(x'|x) \leq \frac{1}{\varepsilon^2}, \quad \|\nabla_{\theta} r_{\theta, \mathbf{Y}}^k(x'|x)\| \leq \frac{2K}{\varepsilon}$$

for $k \geq 1$. Then, (6.3), (6.23) yield

$$\begin{aligned} \|\hat{v}_k^{\theta}(x)\| &\leq \frac{\int r_{\theta, \mathbf{Y}}^k(x|x') |\hat{\xi}_{k-1}^{\theta}|(dx') + \int \|\nabla_{\theta} r_{\theta, \mathbf{Y}}^k(x|x')\| \hat{\xi}_{k-1}^{\theta}(dx')}{\int r_{\theta, \mathbf{Y}}^k(x|x') \hat{\xi}_{k-1}^{\theta}(dx')} \\ &\leq \frac{2K}{\varepsilon^3} + \frac{\|\hat{\xi}_{k-1}^{\theta}\|}{\varepsilon^4} \leq \frac{4\tilde{C}_1 K(1 + \rho_3^{k-1} \|w_{\theta}\|)}{\varepsilon^4} \leq \tilde{C}_2(1 + \rho_4^k \|w_{\theta}\|). \end{aligned}$$

Combining this with (6.4), (6.22), we get

$$(6.24) \quad \|\hat{\alpha}_k^{\theta}(\xi)(B)\| \leq \int_B \|\hat{v}_k^{\theta}(x)\| \xi(dx) \leq \tilde{C}_2(1 + \rho_4^k \|w_{\theta}\|)$$

for $k \geq 0$. Thus, we have

$$(6.25) \quad \|\hat{\alpha}_k^{\theta}(\xi)\| \leq d\tilde{C}_2(1 + \rho_4^k \|w_{\theta}\|) = \tilde{C}_3(1 + \rho_4^k \|w_{\theta}\|).$$

Consequently, Proposition 4.1 implies

$$\begin{aligned} (6.26) \quad \|H_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^{\theta}(\xi))\| &\leq C_2(n - m + \|\hat{\alpha}_m^{\theta}(\xi)\|) \\ &\leq C_2 \tilde{C}_3(1 + n - m + \rho_4^m \|w_{\theta}\|) \\ &\leq \tilde{C}_4(1 + n - m + \rho_4^m \|w_{\theta}\|). \end{aligned}$$

Then, relying on Lemma 6.1 and (6.11), we deduce

$$(6.27) \quad \|\hat{C}_{m:n}^{\theta}\| = \|H_{\theta, \mathbf{Y}}^{m:n}(\hat{\xi}_m^{\theta}, \hat{\alpha}_m^{\theta}(\hat{\xi}_m^{\theta}))\| \leq \tilde{C}_4(1 + n - m + \rho_4^m \|w_{\theta}\|).$$

Moreover, if $m \geq 1$, Proposition 4.1 and (6.7), (6.23), (6.26) yield

$$\begin{aligned} (6.28) \quad \|\hat{H}_{m-1:n}^{\theta}\| &= \|H_{\theta, \mathbf{Y}}^{m-1:n}(\hat{\xi}_{m-1}^{\theta}, \hat{\xi}_{m-1}^{\theta})\| \leq C_2(n - m + \|\hat{\xi}_{m-1}^{\theta}\|) \\ &\leq C_2 \tilde{C}_1(1 + n - m + \rho_4^{m-1} \|w_{\theta}\|) \\ &\leq \tilde{C}_4(1 + n - m + \rho_4^m \|w_{\theta}\|). \end{aligned}$$

The same arguments also imply

$$\begin{aligned} (6.29) \quad \|\hat{H}_{-1:n}^{\theta}\| &= \|H_{\theta, \mathbf{Y}}^{0:n}(\hat{\xi}_{-1}^{\theta}, \hat{\xi}_{-1}^{\theta})\| \leq C_2(n + \|\hat{\xi}_{-1}^{\theta}\|) \leq C_2 \tilde{C}_1(1 + n + \rho_4^{-1} \|w_{\theta}\|) \\ &\leq \tilde{C}_4(1 + n + \rho_4^{-1} \|w_{\theta}\|). \end{aligned}$$

Using (6.26) – (6.29), we conclude that (6.19) holds.

Owing to Proposition 4.1, Lemma 6.1 and (6.5), we have

$$(6.30) \quad \|F_{\theta, \mathbf{Y}}^{m:n}(\xi) - \hat{F}_{m-1:n}^{\theta}\| = \|F_{\theta, \mathbf{Y}}^{m:n}(\xi) - F_{\theta, \mathbf{Y}}^{m:n}(\hat{F}_{m-1:m}^{\theta})\| \leq C_1 \rho_1^{n-m}.$$

Due to the same proposition and (6.25), we also have

$$\begin{aligned}
(6.31) \quad & \|G_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi)) - G_{\theta, \mathbf{Y}}^{m:n}(\xi', \hat{\alpha}_m^\theta(\xi'))\| \leq C_2 \rho_2^{n-m} (1 + \|\hat{\alpha}_m^\theta(\xi)\| + \|\hat{\alpha}_m^\theta(\xi')\|) \\
& \leq 3C_2 \tilde{C}_3 \rho_2^{n-m} (1 + \rho_4^m \|w_\theta\|) \\
& \leq \tilde{C}_4 \rho_2^{n-m} (1 + \rho_4^m \|w_\theta\|).
\end{aligned}$$

Combining Lemma 6.1 and (6.26), (6.30), (6.31), we get

$$\begin{aligned}
(6.32) \quad & \left\| \frac{\xi \hat{\Phi}_{m:n}^\theta}{\langle \xi \hat{R}_{m:n}^\theta \rangle} - \frac{\xi' \hat{\Phi}_{m:n}^\theta}{\langle \xi' \hat{R}_{m:n}^\theta \rangle} \right\| \leq \|G_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi)) - G_{\theta, \mathbf{Y}}^{m:n}(\xi', \hat{\alpha}_m^\theta(\xi'))\| \\
& + \|F_{\theta, \mathbf{Y}}^{m:n}(\xi) - \hat{F}_{m-1:n}^\theta\| \|H_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi))\| \\
& + \|F_{\theta, \mathbf{Y}}^{m:n}(\xi') - \hat{F}_{m-1:n}^\theta\| \|H_{\theta, \mathbf{Y}}^{m:n}(\xi', \hat{\alpha}_m^\theta(\xi'))\| \\
& \leq \tilde{C}_4 \rho_2^{n-m} (1 + \rho_4^m \|w_\theta\|) \\
(6.33) \quad & + 2C_1 \tilde{C}_4 \rho_1^{n-m} (1 + n - m + \rho_4^m \|w_\theta\|) \\
& \leq 3C_1 \tilde{C}_4 \rho_4^{n-m} (1 + n - m + \rho_4^m \|w_\theta\|).
\end{aligned}$$

Similarly, relying on Lemma 6.1 and (6.26), we get

$$\begin{aligned}
(6.34) \quad & \left\| \frac{\xi \hat{\Psi}_{m:n}^\theta}{\langle \xi \hat{R}_{m:n}^\theta \rangle} - \frac{\xi' \hat{\Psi}_{m:n}^\theta}{\langle \xi' \hat{R}_{m:n}^\theta \rangle} \right\| \leq \|H_{\theta, \mathbf{Y}}^{m:n}(\xi, \hat{\alpha}_m^\theta(\xi))\| + \|H_{\theta, \mathbf{Y}}^{m:n}(\xi', \hat{\alpha}_m^\theta(\xi'))\| \\
& \leq 2\tilde{C}_4 (1 + n - m + \rho_4^m \|w_\theta\|).
\end{aligned}$$

Using (6.32), (6.34), we deduce that (6.20) holds. \square

LEMMA 6.3. *Let θ be any element of Θ , while $\varphi : \mathcal{X} \rightarrow [-1, 1]$ is any Borel-measurable function. Moreover, let n, m be any integers satisfying $n \geq m \geq 0$.*

(i) *Suppose that Assumption 2.1 holds. Then, there exists a real number $C_5 \in [1, \infty)$ (independent of $N, \theta, \varphi(x), n, m$ and depending only on ε) such that*

$$(6.35) \quad \left| E \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \mid \mathbf{Y} \right) \right| \leq \frac{C_5 \rho_1^{n-m}}{N},$$

$$(6.36) \quad \left(E \left(\left| \hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right|^2 \mid \mathbf{Y} \right) \right)^{1/2} \leq \frac{C_5 \rho_1^{n-m}}{\sqrt{N}}$$

almost surely (ρ_1 is specified in Proposition 4.1).

(ii) *Suppose that Assumptions 2.1 – 2.3 hold. Then, there exist real numbers $\rho_5 \in (0, 1)$, $C_6 \in [1, \infty)$ (independent of $N, \theta, \varphi(x), n, m$ and depending only on ε, d, K) such that*

$$(6.37) \quad \left\| E \left(\hat{A}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \mid \mathbf{Y} \right) \right\| \leq \frac{C_6(\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N},$$

$$(6.38) \quad \left(E \left(\left\| \hat{A}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \right\|^2 \mid \mathbf{Y} \right) \right)^{1/2} \leq \frac{C_6(\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}},$$

$$(6.39) \quad \left\| E \left(\hat{B}_{m:n}^\theta(\varphi) \mid \mathbf{Y} \right) \right\| \leq \frac{C_6(\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N},$$

$$(6.40) \quad \left(E \left(\left\| \hat{B}_{m:n}^\theta(\varphi) \right\|^2 \mid \mathbf{Y} \right) \right)^{1/2} \leq \frac{C_6(\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}}$$

almost surely.

Proof. (i) Throughout this part of the proof, the following notation is used. ξ, ξ' are any elements of $\mathcal{P}(\mathcal{X})$. $\mathbf{1}(x)$ is the function which maps $x \in \mathcal{X}$ to one.

Relying on (2.1), (2.2), (2.10), (4.2), (4.4), (6.5), (6.8), we conclude

(6.41)

$$\begin{aligned} P\left(\hat{X}_{k,1}^\theta \in B_1, \dots, \hat{X}_{k,N}^\theta \in B_N \mid \mathbf{Y}, \hat{\xi}_{k-1}^\theta\right) &= \prod_{i=1}^N \left(\frac{\int \left(\int_{B_i}(x') r_{\theta, \mathbf{Y}}^k(x'|x) \mu(dx') \right) \hat{\xi}_{k-1}^\theta(dx)}{\int \left(\int r_{\theta, \mathbf{Y}}^k(x'|x) \mu(dx') \right) \hat{\xi}_{k-1}^\theta(dx)} \right) \\ &= \prod_{i=1}^N \frac{(\hat{\xi}_{k-1}^\theta \hat{R}_{k-1:k}^\theta)(B_i)}{\langle \hat{\xi}_{k-1}^\theta \hat{R}_{k-1:k}^\theta \rangle} = \prod_{i=1}^N \hat{F}_{n-1:n}^\theta(B_i) \end{aligned}$$

almost surely for any $B_1, \dots, B_N \in \mathcal{B}(\mathcal{X})$, $k \geq 1$. Similarly, using (6.2), (6.5), we deduce

$$(6.42) \quad P\left(\hat{X}_{0,1}^\theta \in B_1, \dots, \hat{X}_{0,N}^\theta \in B_N \mid \mathbf{Y}, \hat{\xi}_{-1}^\theta\right) = \prod_{i=1}^N \hat{\xi}_{-1}^\theta(B_i) = \prod_{i=1}^N \hat{F}_{-1:0}^\theta(B_i)$$

almost surely for the same B_1, \dots, B_N . Moreover, Lemma 6.1 and (4.4), (6.5), (6.8) imply

$$(6.43) \quad \hat{F}_{m-1:n}^\theta(\varphi) = \frac{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})}, \quad \hat{F}_{m:n}^\theta(\varphi) = \frac{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})}.$$

Let $C_5 = 2C_1^3$ (C_1 is specified in Proposition 4.1). Owing to Proposition 4.1 and (4.4), (6.5), (6.8), we have

$$(6.44) \quad \left| \frac{(\xi \hat{R}_{m:n}^\theta)(\varphi)}{(\xi \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\xi' \hat{R}_{m:n}^\theta)(\varphi)}{(\xi' \hat{R}_{m:n}^\theta)(\mathbf{1})} \right| = |F_{\theta, \mathbf{Y}}^{m:n}(\xi)(\varphi) - F_{\theta, \mathbf{Y}}^{m:n}(\xi')(\varphi)| \leq C_1 \rho_1^{n-m}.$$

Due to the same arguments, we also have

$$(6.45) \quad \frac{(\xi \hat{R}_{m:n}^\theta)(\mathbf{1})}{(\xi' \hat{R}_{m:n}^\theta)(\mathbf{1})} \leq C_1.$$

Using Proposition 3.1 and (6.41) – (6.45), we conclude

$$\begin{aligned} (6.46) \quad & \left| E\left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \mid \mathbf{Y}, \hat{\xi}_{m-1}^\theta\right) \right| \\ &= \left| E\left(\frac{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} \mid \mathbf{Y}, \hat{\xi}_{m-1}^\theta\right) \right| \\ &\leq \frac{2C_1^3 \rho_1^{n-m}}{N} = \frac{C_5 \rho_1^{n-m}}{N} \end{aligned}$$

almost surely.⁴ Relying on the same arguments, we deduce

$$\begin{aligned}
(6.47) \quad & E \left(\left| \hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
&= E \left(\left| \frac{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} \right|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
&\leq \left(\frac{2C_1^2 \rho_1^{n-m}}{\sqrt{N}} \right)^2 \leq \left(\frac{C_5 \rho_1^{n-m}}{\sqrt{N}} \right)^2
\end{aligned}$$

almost surely. Combining (6.46), (6.47) with the tower property of conditional expectations, we conclude that (6.35), (6.36) hold almost surely.

(ii) Let $\xi, \xi', \mathbf{1}(x)$ have the same meaning as in (i). Using (6.10), (6.11), (6.13), it is straightforward to verify

$$(6.48) \quad \hat{A}_{m:n}^\theta(\varphi) = \frac{(\hat{\xi}_m^\theta \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})}, \quad \hat{C}_{m:n}^\theta(\varphi) = \frac{(\hat{\xi}_m^\theta \hat{\Psi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})}.$$

Similarly, Lemma 6.1 yields

$$(6.49) \quad \hat{G}_{m-1:n}^\theta(\varphi) = \frac{(\hat{F}_{m-1:m}^\theta \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})}, \quad \hat{H}_{m-1:n}^\theta(\varphi) = \frac{(\hat{F}_{m-1:m}^\theta \hat{\Psi}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})}.$$

Let $\rho_5 = \max\{\sqrt{\rho_1}, \sqrt{\rho_4}\}$, $\tilde{C}_1 = \max_{n \geq 1} n \rho_5^n$, $\tilde{C}_2 = 2C_4 \tilde{C}_1$, $\tilde{C}_3 = 2C_1^2 \tilde{C}_2$ (ρ_1, ρ_4, C_1, C_4 are specified in Proposition 4.1 and Lemma 6.2). Since $\rho_4 \leq \rho_5$, $\rho_4^{n-m}(n-m) \leq \tilde{C}_1 \rho_5^{n-m}$, Lemma 6.2 implies

$$\begin{aligned}
(6.50) \quad & \left\| \frac{(\xi \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\xi \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\xi' \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\xi' \hat{R}_{m:n}^\theta)(\mathbf{1})} \right\| \leq C_4 \rho_4^{n-m} (1 + n - m + \rho_4^m \|w_\theta\|) \\
&\leq 2C_4 \tilde{C}_1 (\rho_5^{n-m} + \rho_5^m \|w_\theta\|) \\
&\leq \tilde{C}_2 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|).
\end{aligned}$$

Using Proposition 3.1 and (6.41), (6.42), (6.45), (6.48) – (6.50), we conclude

$$\begin{aligned}
(6.51) \quad & \left\| E \left(\hat{A}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right\| \\
&= \left\| E \left(\frac{(\hat{\xi}_m^\theta \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\hat{F}_{m-1:m}^\theta \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right\| \\
&\leq \frac{2C_1^2 \tilde{C}_2 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N} = \frac{\tilde{C}_3 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N}
\end{aligned}$$

⁴To get (6.46), (6.47), the following should be done: In Proposition 3.1, set $z = x$, $k = N$ and replace $f(z)$, $g(z)$, $\xi_k(dz)$, $\xi(dz)$ with $(\hat{R}_{m:n}^\theta \varphi)(x)$, $(\hat{R}_{m:n}^\theta \mathbf{1})(x)$, $\hat{\xi}_m^\theta(dx)$, $\hat{F}_{m-1:m}^\theta(dx)$.

almost surely.⁵ Relying on the same arguments, we deduce

$$\begin{aligned}
(6.52) \quad & E \left(\left\| \hat{A}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
&= E \left(\left\| \frac{(\hat{\xi}_m^\theta \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\hat{F}_{m-1:m}^\theta \hat{\Phi}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
&\leq \left(\frac{2C_1 \tilde{C}_2 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}} \right)^2 \leq \left(\frac{\tilde{C}_3 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}} \right)^2
\end{aligned}$$

almost surely.

Let $\tilde{C}_4 = C_5 \tilde{C}_2$ (C_5 is defined in (i)). Since $\tilde{C}_1 \rho_5^{m-n} \geq n - m$, Lemma 6.2 implies

$$\begin{aligned}
(6.53) \quad & \left\| \hat{H}_{m-1:n}^\theta(\varphi) \right\| \leq C_4 (1 + n - m + \rho_4^m \|w_\theta\|) \leq 2C_4 \tilde{C}_1 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|) \\
&\leq \tilde{C}_2 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|).
\end{aligned}$$

As $\hat{H}_{m-1:n}^\theta(\mathbf{1})$ is measurable with respect to \mathbf{Y} , $\hat{\xi}_{m-1}^\theta$ and $\rho_1^{n-m} \leq \rho_5^{2(n-m)}$, (6.46), (6.53) yield

$$\begin{aligned}
(6.54) \quad & \left\| E \left(\left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right) \hat{H}_{m-1:n}^\theta(\mathbf{1}) \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right\| \\
&\leq \left| E \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right| \left\| \hat{H}_{m-1:n}^\theta(\mathbf{1}) \right\| \\
&\leq \frac{C_5 \tilde{C}_2 \rho_1^{n-m} (\rho_5^{m-n} + \rho_5^m \|w_\theta\|)}{N} \leq \frac{\tilde{C}_4 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N}
\end{aligned}$$

almost surely.

Let $\tilde{C}_5 = C_5 \tilde{C}_3$ (C_5 is defined in (i)). Since $\tilde{C}_1 \rho_5^{m-n} \geq n - m$, Lemma 6.2 implies

$$\begin{aligned}
(6.55) \quad & \left\| \hat{C}_{m:n}^\theta(\varphi) \right\| \leq C_4 (1 + n - m + \rho_4^m \|w_\theta\|) \leq 2C_4 \tilde{C}_1 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|) \\
&\leq \tilde{C}_2 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|).
\end{aligned}$$

Moreover, the same lemma yields

$$\begin{aligned}
(6.56) \quad & \left\| \frac{(\hat{\xi}_m^\theta \hat{\Psi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\hat{\xi}' \hat{\Psi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}' \hat{R}_{m:n}^\theta)(\mathbf{1})} \right\| \leq C_4 (1 + n - m + \rho_4^m \|w_\theta\|) \\
&\leq \tilde{C}_2 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|).
\end{aligned}$$

Using Proposition 3.1 and (6.41), (6.42), (6.45), (6.49), (6.48), (6.56), we conclude

$$\begin{aligned}
(6.57) \quad & E \left(\left\| \hat{C}_{m:n}^\theta(\varphi) - \hat{H}_{m-1:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
&= E \left(\left\| \frac{(\hat{\xi}_m^\theta \hat{\Psi}_{m:n}^\theta)(\varphi)}{(\hat{\xi}_m^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} - \frac{(\hat{F}_{m-1:m}^\theta \hat{\Psi}_{m:n}^\theta)(\varphi)}{(\hat{F}_{m-1:m}^\theta \hat{R}_{m:n}^\theta)(\mathbf{1})} \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
&\leq \left(\frac{2C_1 \tilde{C}_2 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|)}{\sqrt{N}} \right)^2 \leq \left(\frac{\tilde{C}_3 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|)}{\sqrt{N}} \right)^2
\end{aligned}$$

⁵To get (6.51), (6.52), the following should be done: In Proposition 3.1, set $z = x$, $k = N$ and replace $f(z)$, $g(z)$, $\xi_k(dz)$, $\xi(dz)$ with $(\hat{\Phi}_{m:n}^\theta \varphi)(x)$, $(\hat{R}_{m:n}^\theta \mathbf{1})(x)$, $\hat{\xi}_m^\theta(dx)$, $\hat{F}_{m-1:m}^\theta(dx)$.

almost surely.⁶ As $\rho_1^{n-m} \leq \rho_5^{2(n-m)}$, Hölder inequality and (6.47), (6.57) imply

$$\begin{aligned}
(6.58) \quad & \left\| E \left(\left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right) \left(\hat{C}_{m:n}^\theta(\mathbf{1}) - \hat{H}_{m:n}^\theta(\mathbf{1}) \right) \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right\| \\
& \leq \left(E \left(\left| \hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right)^{1/2} \\
& \quad \cdot \left(E \left(\left\| \hat{C}_{m:n}^\theta(\mathbf{1}) - \hat{H}_{m:n}^\theta(\mathbf{1}) \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right)^{1/2} \\
& \leq \frac{C_5 \tilde{C}_3 \rho_1^{n-m} (\rho_5^{m-n} + \rho_5^m \|w_\theta\|)}{N} \leq \frac{\tilde{C}_5 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N}
\end{aligned}$$

almost surely. Similarly, (6.47), (6.55) yield

$$\begin{aligned}
(6.59) \quad & E \left(\left\| \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right) \hat{C}_{m:n}^\theta(\mathbf{1}) \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
& \leq \tilde{C}_2^2 (\rho_5^{m-n} + \rho_5^m \|w_\theta\|)^2 E \left(\left| \hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \\
& \leq \left(\frac{C_5 \tilde{C}_2 \rho_1^{n-m} (\rho_5^{m-n} + \rho_5^m \|w_\theta\|)}{\sqrt{N}} \right)^2 \leq \left(\frac{\tilde{C}_4 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}} \right)^2
\end{aligned}$$

almost surely.

Let $\tilde{C}_6 = \tilde{C}_4 + \tilde{C}_5$. Due to (6.12), we have

$$\begin{aligned}
(6.60) \quad & \hat{B}_{m:n}^\theta(\varphi) = - \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right) \hat{C}_{m:n}^\theta(\mathbf{1}) \\
& = - \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right) \left(\hat{C}_{m:n}^\theta(\mathbf{1}) - \hat{H}_{m:n}^\theta(\mathbf{1}) \right) \\
& \quad - \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right) \hat{H}_{m:n}^\theta(\mathbf{1}).
\end{aligned}$$

Consequently, (6.54), (6.58) and the second part of (6.60) imply

$$(6.61) \quad \left\| E \left(\hat{B}_{m:n}^\theta(\varphi) \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \right\| \leq \frac{(\tilde{C}_4 + \tilde{C}_5) (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N} = \frac{\tilde{C}_6 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N}$$

almost surely. Moreover, the first part of (6.60) and (6.59) yield

$$(6.62) \quad E \left(\left\| \hat{B}_{m:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y}, \hat{\xi}_{m-1}^\theta \right) \leq \left(\frac{\tilde{C}_4 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}} \right)^2 \leq \left(\frac{\tilde{C}_6 (\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}} \right)^2$$

almost surely.

Let $C_6 = \tilde{C}_3 + \tilde{C}_6$. Then, combining (6.51), (6.52), (6.61), (6.62) with the tower property of conditional expectations, we conclude that (6.37) – (6.40) hold almost surely. \square

⁶To get (6.57), the following should be done: In Proposition 3.1, set $z = x$, $k = N$ and replace $f(z)$, $g(z)$, $\xi_k(dz)$, $\xi(dz)$ with $(\hat{\Psi}_{m:n}^\theta \varphi)(x)$, $(\hat{R}_{m:n}^\theta \mathbf{1})(x)$, $\hat{\xi}_m^\theta(dx)$, $\hat{F}_{m-1:m}^\theta(dx)$.

PROPOSITION 6.4. Let θ be any element of Θ , while $\mathbf{y} = \{y_n\}_{n \geq 0}$ is any sequence in \mathcal{Y} . Moreover, let $\varphi : \mathcal{X} \rightarrow [-1, 1]$ be any Borel-measurable function, while n is any non-negative integer.

(i) Suppose that Assumption 2.1 holds. Then, there exists a real number $L \in [1, \infty)$ (independent of N , θ , \mathbf{y} , $\varphi(x)$, n and depending only on ε) such that

$$(6.63) \quad \left| E \left(\hat{\xi}_n^\theta(\varphi) - F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(\varphi) \mid \mathbf{Y} = \mathbf{y} \right) \right| \leq \frac{L}{N},$$

$$(6.64) \quad \left(E \left(\left| \hat{\xi}_n^\theta(\varphi) - F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(\varphi) \right|^2 \mid \mathbf{Y} = \mathbf{y} \right) \right)^{1/2} \leq \frac{L}{\sqrt{N}}.$$

(ii) Suppose that Assumptions 2.1 – 2.3 hold. Then, there exist real numbers $\rho \in (0, 1)$, $M \in [1, \infty)$ (independent of N , θ , \mathbf{y} , $\varphi(x)$, n and depending only on ε , d , K) such that

$$(6.65) \quad \left\| E \left(\hat{\zeta}_n^\theta(\varphi) - G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(\varphi) \mid \mathbf{Y} = \mathbf{y} \right) \right\| \leq \frac{M(1 + \rho^n \|w_\theta\|)}{N},$$

$$(6.66) \quad \left(E \left(\left\| \hat{\zeta}_n^\theta(\varphi) - G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(\varphi) \right\|^2 \mid \mathbf{Y} = \mathbf{y} \right) \right)^{1/2} \leq \frac{M(1 + \rho^n \|w_\theta\|)}{\sqrt{N}}.$$

REMARK. Relying on (2.3), (2.4), (4.4), (4.6), it is easy to show

$$P_{\theta, \mathbf{Y}}^n(B) = F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(B), \quad Q_{\theta, \mathbf{Y}}^n(B) = G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(B)$$

for $B \in \mathcal{B}(\mathcal{X})$, $n \geq 1$. Hence, Proposition 6.4 can be considered as an extended version of Theorem 2.1. Moreover, the bounds in (6.64), (6.66) can be viewed as by-products of Theorem 2.1. Under the same conditions as in Proposition 6.4, bounds similar to (6.64), (6.66) have been derived in [11].

Proof. (i) Let $L = C_5(1 - \rho_1)^{-1}$ (ρ_1 is specified in Proposition 4.1). Using (6.2), (6.5), it is straightforward to verify

$$\hat{\xi}_n^\theta(\varphi) = \hat{F}_{n:n}^\theta(\varphi), \quad F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(\varphi) = F_{\theta, \mathbf{Y}}^{0:n}(\hat{\xi}_{-1}^\theta)(\varphi) = \hat{F}_{-1:n}^\theta(\varphi).$$

Therefore, we get

$$(6.67) \quad \hat{\xi}_n^\theta(\varphi) - F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(\varphi) = \hat{F}_{n:n}^\theta(\varphi) - \hat{F}_{-1:n}^\theta(\varphi) = \sum_{m=0}^n \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right).$$

Then, Lemma 6.3 implies

$$(6.68) \quad \left| E \left(\hat{\xi}_n^\theta(\varphi) - F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(\varphi) \mid \mathbf{Y} \right) \right| \leq \sum_{m=0}^n \left| E \left(\hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \mid \mathbf{Y} \right) \right| \\ \leq \frac{C_5}{N} \sum_{m=0}^n \rho_1^{n-m} \leq \frac{L}{N}$$

almost surely. Moreover, Minkowski inequality, Lemma 6.3 and (6.67) yield

$$(6.69) \quad \left(E \left(\left| \hat{\xi}_n^\theta(\varphi) - F_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta)(\varphi) \right|^2 \mid \mathbf{Y} \right) \right)^{1/2} \leq \sum_{m=0}^n \left(E \left(\left| \hat{F}_{m:n}^\theta(\varphi) - \hat{F}_{m-1:n}^\theta(\varphi) \right|^2 \mid \mathbf{Y} \right) \right)^{1/2} \\ \leq \frac{C_5}{\sqrt{N}} \sum_{m=0}^n \rho_1^{n-m} \leq \frac{L}{\sqrt{N}}$$

almost surely. Using (6.68), (6.69), we conclude that (6.63), (6.64) hold.

(ii) Let $\rho = \sqrt{\rho_5}$, $\tilde{C} = \max_{n \geq 1} n\rho^n$, while $M = 4C_6\tilde{C}(1-\rho)^{-1}$ (ρ_5, C_6 are specified in Lemma 6.3). Owing to Lemmas 6.1, 6.3, we have

$$(6.70) \quad \begin{aligned} \left\| E \left(\hat{G}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \middle| \mathbf{Y} \right) \right\| &\leq \left\| E \left(\hat{A}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \middle| \mathbf{Y} \right) \right\| \\ &\quad + \left\| E \left(\hat{B}_{m:n}^\theta(\varphi) \middle| \mathbf{Y} \right) \right\| \\ &\leq \frac{2C_6(\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{N} \end{aligned}$$

almost surely. Similarly, due to Minkowski inequality and Lemmas 6.1, 6.3, we have

$$(6.71) \quad \begin{aligned} \left(E \left(\left\| \hat{G}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y} \right) \right)^{1/2} &\leq \left(E \left(\left\| \hat{A}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y} \right) \right)^{1/2} \\ &\quad + \left(E \left(\left\| \hat{B}_{m:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y} \right) \right)^{1/2} \\ &\leq \frac{2C_6(\rho_5^{n-m} + \rho_5^n \|w_\theta\|)}{\sqrt{N}} \end{aligned}$$

almost surely.

Using (6.2), (6.6), it is straightforward to verify

$$\hat{\zeta}_n^\theta(\varphi) = \hat{G}_{n:n}^\theta(\varphi), \quad G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(\varphi) = G_{\theta, \mathbf{Y}}^{0:n}(\hat{\xi}_{-1}^\theta, \hat{\zeta}_{-1}^\theta)(\varphi) = \hat{G}_{-1:n}^\theta(\varphi).$$

Therefore, we get

$$(6.72) \quad \hat{\zeta}_n^\theta(\varphi) - G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(\varphi) = \hat{G}_{n:n}^\theta(\varphi) - G_{-1:n}^\theta(\varphi) = \sum_{m=0}^n \left(\hat{G}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \right).$$

Then, (6.70) implies

$$(6.73) \quad \begin{aligned} \left\| E \left(\hat{\zeta}_n^\theta(\varphi) - G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(\varphi) \middle| \mathbf{Y} \right) \right\| &\leq \sum_{m=0}^n \left\| E \left(\hat{G}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \middle| \mathbf{Y} \right) \right\| \\ &\leq \frac{2C_6}{N} \sum_{m=0}^n (\rho_5^{n-m} + \rho_5^n \|w_\theta\|) \leq \frac{2C_6}{(1-\rho_5)N} + \frac{2C_6(n+1)\rho_5^n \|w_\theta\|}{N} \leq \frac{M(1+\rho^n \|w_\theta\|)}{N} \end{aligned}$$

almost surely. Moreover, Minkowski inequality and (6.71) – (6.72) yield

$$(6.74) \quad \begin{aligned} \left(E \left(\left\| \hat{\zeta}_n^\theta(\varphi) - G_{\theta, \mathbf{Y}}^{0:n}(\xi_\theta, \zeta_\theta)(\varphi) \right\|^2 \middle| \mathbf{Y} \right) \right)^{1/2} &\leq \sum_{m=0}^n \left(E \left(\left\| \hat{G}_{m:n}^\theta(\varphi) - \hat{G}_{m-1:n}^\theta(\varphi) \right\|^2 \middle| \mathbf{Y} \right) \right)^{1/2} \\ &\leq \frac{2C_6}{\sqrt{N}} \sum_{m=0}^n (\rho_5^{n-m} + \rho_5^n \|w_\theta\|) \leq \frac{2C_6}{(1-\rho_5)\sqrt{N}} + \frac{2C_6(n+1)\rho_5^n \|w_\theta\|}{\sqrt{N}} \leq \frac{M(1+\rho^n \|w_\theta\|)}{\sqrt{N}} \end{aligned}$$

almost surely. Using (6.73), (6.74), we conclude that (6.65), (6.66) hold. \square

REFERENCES

- [1] C. Andrieu, A. Doucet, S. S. Singh, and V. B. Tadić, *Particle methods for change detection, system identification, and control*, Proceedings of IEEE, 92 (2004), pp. 423–438.
- [2] P. Brémaud, *Discrete Probability Models and Methods*, Springer-Verlag, 2017.
- [3] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer-Verlag, 2005.
- [4] D. Crisan and B. Rozovskii (Eds.), *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, 2011.
- [5] P. Del Moral and A. Guionnet, *On the stability of interacting processes and with applications to filtering and genetic algorithms*, Annales de l’Institut Henri Poincaré (B) Probability and Statistics, 37 (2001), pp. 155–194.
- [6] P. Del Moral, *Feynman-Kac Formulae*, Springer-Verlag, 2004.
- [7] P. Del Moral, *Mean Field Simulation for Monte Carlo Integration*, CRC Press, 2013.
- [8] P. Del Moral, A. Doucet, and G. W. Peters, *Sharp propagation of chaos estimates for Feynman-Kac particle models*, Theory of Probability and Its Applications, 51 (2007), pp. 459–485.
- [9] P. Del Moral, A. Doucet, and S. S. Singh, *A backward interpretation of Feynman-Kac formulae*, ESAIM: Mathematical Modelling and Numerical Analysis, 44 (2010), pp. 947–975.
- [10] P. Del Moral, P. Jacob, A. Lee, L. Murray, and G. W. Peters, *Feynman-Kac particle integration with geometric interacting jumps*, Stochastic Analysis and Applications, 31 (2013), pp. 830–871.
- [11] P. Del Moral, A. Doucet, and S. S. Singh, *Uniform stability of a particle approximation of the optimal filter derivative*, SIAM Journal on Control and Optimization, 53 (2015), pp. 1278–1304.
- [12] P. Del Moral and A. Jasra, *A sharp first order analysis of Feynman-Kac particle models, Part I: Propagation of chaos*, Stochastic Processes and their Applications, 128 (2018), pp. 332–353.
- [13] R. Douc, E. Moulines, and T. Ryden, *Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime*, Annals of Statistics, 32 (2004), pp. 2254–2304.
- [14] R. Douc, E. Moulines, and D. S. Stoffer, *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*, CRC Press, 2014.
- [15] A. Doucet, N. de Freitas and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.
- [16] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, *On particle methods for parameter estimation in state-space models*, Statistical Science, 30 (2015), pp. 328–351.
- [17] F. Le Gland and L. Mével, *Recursive estimation in hidden Markov models*, Proceedings of the 36th Conference on Decision and Control, pp. 3468–3473, 1997.
- [18] F. Le Gland and L. Mével, *Exponential forgetting and geometric ergodicity in hidden Markov models*, Mathematics of Control, Signals and Systems 13 (2000), pp. 63–93.
- [19] F. Le Gland and N. Oudjane, *Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters*, Annals of Applied Probability, 14 (2004), pp. 144–187.
- [20] J. Olsson and J. Westerborn Alenlöv, *Particle-based, online estimation of tangent filters with application to parameter estimation in nonlinear state-space models*, Annals of the Institute of Statistical Mathematics, accepted for publication.
- [21] N. Oudjane and S. Rubenthaler, *Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals*, Stochastic Analysis and Applications, 23 (2005), pp. 421–448.
- [22] G. Poyiadjis, A. Doucet, and S. S. Singh, *Particle approximations of the score and observed information matrix in state space models with application to parameter estimation*, Biometrika, 98 (2011), pp. 65–80.
- [23] V. B. Tadić and A. Doucet, *Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models*, Stochastic Processes and Their Applications, 115 (2005), pp. 1408–1436.
- [24] V. B. Tadić and A. Doucet, *Asymptotic bias of stochastic gradient search*, Annals of Applied Probability, 27 (2017), pp. 3255–3304.
- [25] V. Z. B. Tadić and A. Doucet, *Asymptotic properties of recursive maximum likelihood estimation in non-linear state-space models*, available at arXiv:1806.09571.