

# Stability of Optimal Filter Higher-Order Derivatives

Vladislav Z. B. Tadić \*

Arnaud Doucet†

**Abstract.** In many scenarios, a state-space model depends on a parameter which needs to be inferred from data. Using stochastic gradient search and the optimal filter first-order derivatives, the parameter can be estimated online. To analyze the asymptotic behavior of such methods, it is necessary to establish results on the existence and stability of the optimal filter higher-order derivatives. These properties are studied here. Under regularity conditions, we show that the optimal filter higher-order derivatives exist and forget initial conditions exponentially fast. We also show that the same derivatives are geometrically ergodic.

**Keywords.** State-Space Models, Optimal Filter, Optimal Filter Higher-Order Derivatives, Forgetting of Initial Conditions, Geometric Ergodicity, Log-Likelihood.

**AMS Subject Classification.** Primary 60G35; Secondary 62M20, 93E11.

## 1. Introduction

State-space models, also known as continuous-state hidden Markov models, are a powerful and versatile tool for statistical modeling of complex time-series data and stochastic dynamic systems. These models can be viewed as a discrete-time Markov process which are observed only through noisy measurements of their states. In this context, one of the most important problems is the optimal estimation of the current state given the noisy measurements of the current and previous states. This problem is known as optimal filtering. Optimal filtering has been studied in a number of papers and books; see, e.g., [3], [4], [9] and references therein.

In many applications, a state-space model depends on a parameter whose value needs to be inferred from data. When the number of data points is large, it is desirable, for the sake of computational efficiency, to infer the parameter recursively (i.e., online). In the maximum likelihood approach, recursive parameter estimation can be performed using stochastic gradient search, where the underlying gradient estimation is based on the optimal filter and its first-order derivatives; see, e.g., [10], [15], [17]. In [17], it has been shown that the asymptotic behavior of recursive maximum likelihood estimation in finite-state hidden Markov models is closely related to the analytical properties, higher-order differentiability and analyticity, of the underlying log-likelihood rate. In view of the recent results on stochastic gradient search [19], a similar relationship is likely to hold for state-space models. However, to apply the results of [19] to recursive maximum likelihood estimation in state-space models, it is necessary to establish results on the higher-order differentiability of the log-likelihood rate for these models. Since the log-likelihood rate for state-space models is a functional of the optimal filter, the analytical properties of this rate are tightly connected to the existence and stability of the optimal filter higher-order derivatives. Hence, one of the first steps to carry out asymptotic analysis of recursive maximum likelihood estimation in state-space models is to establish results on the existence and stability of these derivatives. To the best of our knowledge, this problem has never been addressed before and the results presented here fill this gap in the literature on optimal filtering.

In this paper, the optimal filter higher-order derivatives and their existence and stability properties are studied. Under standard stability and regularity conditions, we show that these derivatives exist and forget initial conditions exponentially fast. We also show that the optimal filter higher-order derivatives

---

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

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

are geometrically ergodic. The obtained results cover state-space models met in practice and are one of the first stepping stones to analyze the asymptotic behavior of recursive maximum likelihood estimation in non-linear state-space models [20].

The paper is organized as follows. In Section 2, the existence and stability of the optimal filter higher-order derivatives are studied and the main results are presented. In Section 3, the main results are used to study the analytical properties of log-likelihood for state-space models. An example illustrating the main results is provided in Section 4. In Sections 5 – 8, the main results and their corollaries are proved.

## 2. Main Results

### 2.1. State-Space Models and Optimal Filter

To specify state-space models and to formulate the problem of optimal filtering, we use the following notation. For a set  $\mathcal{Z}$  in a metric 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  $\mathcal{Y} \in \mathcal{B}(\mathbb{R}^{d_y})$ .  $P(x, dx')$  is a transition kernel on  $\mathcal{X}$ , while  $Q(x, dy)$  is a conditional probability measure on  $\mathcal{Y}$  given  $x \in \mathcal{X}$ .  $(\Omega, \mathcal{F}, P)$  is a probability space. Then, a state-space model can be defined as an  $\mathcal{X} \times \mathcal{Y}$ -valued stochastic process  $\{(X_n, Y_n)\}_{n \geq 0}$  on  $(\Omega, \mathcal{F}, P)$  which satisfies

$$P((X_{n+1}, Y_{n+1}) \in B | X_{0:n}, Y_{0:n}) = \int I_B(x, y) Q(x, dy) P(X_n, dx)$$

almost surely for any  $B \in \mathcal{B}(\mathcal{X} \times \mathcal{Y})$  and  $n \geq 0$ .  $\{X_n\}_{n \geq 0}$  are the unobservable states, while  $\{Y_n\}_{n \geq 0}$  are the observations. One of the most important problems related to state-space models is the estimation of the current state  $X_n$  given the state-observations  $Y_{1:n}$ . This problem is known as filtering. In the Bayesian approach, the optimal estimation of  $X_n$  given  $Y_{1:n}$  is based on the (optimal) filtering distribution  $P(X_n \in dx_n | Y_{1:n})$ . As  $P(x, dx')$  and  $Q(x, dy)$  are rarely available in practice, the filtering distribution is usually computed 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 state-space models. To define such a family, we rely on the following notation. Let  $d \geq 1$  be an integer, while  $\Theta \subset \mathbb{R}^d$  is a bounded open set.  $\mathcal{P}(\mathcal{X})$  is the set of probability measures on  $\mathcal{X}$ , while  $\mu(dx)$  and  $\nu(dy)$  are measures on  $\mathcal{X}$  and  $\mathcal{Y}$  (respectively).  $p_\theta(x'|x)$  and  $q_\theta(y|x)$  are functions which map  $\theta \in \Theta$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$  to  $[0, \infty)$  and satisfy

$$\int p_\theta(x'|x) \mu(dx') = \int q_\theta(y|x) \nu(dy) = 1$$

for all  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ . With this notation, a parametric family of state-space models can be defined as an  $\mathcal{X} \times \mathcal{Y}$ -valued stochastic process  $\{(X_n^{\theta, \lambda}, Y_n^{\theta, \lambda})\}_{n \geq 0}$  on  $(\Omega, \mathcal{F}, P)$  which is parameterized by  $\theta \in \Theta$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$  and satisfies

$$\begin{aligned} P((X_0^{\theta, \lambda}, Y_0^{\theta, \lambda}) \in B) &= \int \int I_B(x, y) q_\theta(y|x) \lambda(dx) \nu(dy), \\ P((X_{n+1}^{\theta, \lambda}, Y_{n+1}^{\theta, \lambda}) \in B | X_{0:n}^{\theta, \lambda}, Y_{0:n}^{\theta, \lambda}) &= \int \int I_B(x, y) q_\theta(y|x) p_\theta(x | X_n^{\theta, \lambda}) \mu(dx) \nu(dy), \end{aligned}$$

almost surely for any  $B \in \mathcal{B}(\mathcal{X} \times \mathcal{Y})$  and  $n \geq 0$ .<sup>1</sup>

To show how the filtering distribution is computed using approximate model  $\{(X_n^{\theta, \lambda}, Y_n^{\theta, \lambda})\}_{n \geq 0}$ , we use the following notation.  $r_\theta(y, x'|x)$  is the function defined by

$$r_\theta(y, x'|x) = q_\theta(y|x') p_\theta(x'|x) \tag{1}$$

---

<sup>1</sup>To evaluate the values of  $\theta$  for which  $\{(X_n^{\theta, \lambda}, Y_n^{\theta, \lambda})\}_{n \geq 0}$  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], [9] and references cited therein.

for  $\theta \in \Theta$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ , while  $r_{\theta, \mathbf{y}}^{m:n}(x'|x)$  is the function recursively defined by

$$r_{\theta, \mathbf{y}}^{m:m+1}(x'|x) = r_{\theta}(y_{m+1}, x'|x), \quad r_{\theta, \mathbf{y}}^{m:n+1}(x'|x) = \int r_{\theta}(y_{n+1}, x'|x'') r_{\theta, \mathbf{y}}^{m:n}(x''|x) \mu(dx'') \quad (2)$$

for  $n > m \geq 0$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ .  $p_{\theta, \mathbf{y}}^{m:n}(x|\lambda)$  and  $P_{\theta, \mathbf{y}}^{m:n}(dx|\lambda)$  are the function and the probability measure defined by

$$p_{\theta, \mathbf{y}}^{m:n}(x|\lambda) = \frac{\int r_{\theta, \mathbf{y}}^{m:n}(x|x') \lambda(dx')}{\int \int r_{\theta, \mathbf{y}}^{m:n}(x''|x') \mu(dx'') \lambda(dx')}, \quad P_{\theta, \mathbf{y}}^{m:n}(B|\lambda) = \int_B p_{\theta, \mathbf{y}}^{m:n}(x'|\lambda) \mu(dx') \quad (3)$$

for  $B \in \mathcal{B}(\mathcal{X})$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ , while  $P_{\theta, \mathbf{y}}^{m:n}(\lambda)$  is a ‘short-hand’ notation for  $P_{\theta, \mathbf{y}}^{m:n}(dx|\lambda)$ . Then, it can easily be shown that  $P_{\theta, \mathbf{y}}^{m:n}(\lambda)$  is the filtering distribution, i.e.,

$$P_{\theta, \mathbf{y}}^{0:n}(B|\lambda) = P\left(X_n^{\theta, \lambda} \in B \mid Y_{1:n}^{\theta, \lambda} = y_{1:n}\right)$$

for each  $\theta \in \Theta$ ,  $B \in \mathcal{B}(\mathcal{X})$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $n \geq 1$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ . In this context,  $\lambda$  can be interpreted as the initial condition of the filtering distribution  $P_{\theta, \mathbf{y}}^{m:n}(\lambda)$ .

## 2.2. Optimal Filter Higher-Order Derivatives

Let  $p \geq 1$  be an integer. Throughout the paper, we assume that  $p_{\theta}(x'|x)$  and  $q_{\theta}(y|x)$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \Theta$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ .

To define the higher-order derivatives of the optimal filter, we use the following notation.  $\mathbb{N}_0$  is the set of non-negative integers.  $\mathbf{0}$  is the element of  $\mathbb{N}_0^d$  whose all components are zero. For  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d) \in \mathbb{N}_0^d$ , relation  $\boldsymbol{\beta} \leq \boldsymbol{\alpha}$  is taken component-wise, i.e.,  $\boldsymbol{\beta} \leq \boldsymbol{\alpha}$  if and only if  $\alpha_i \leq \beta_i$  for each  $1 \leq i \leq d$ . For the same  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  satisfying  $\boldsymbol{\beta} \leq \boldsymbol{\alpha}$ ,  $\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}}$  denotes the multinomial coefficient

$$\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_d}{\beta_d}.$$

For  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \Theta$ , notation  $|\boldsymbol{\alpha}|$  and  $\partial_{\boldsymbol{\theta}}^{\boldsymbol{\alpha}}$  stand for

$$|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d, \quad \partial_{\boldsymbol{\theta}}^{\boldsymbol{\alpha}} = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial \theta_1^{\alpha_1} \cdots \partial \theta_d^{\alpha_d}}.$$

$d(p)$  is the number elements in set  $\{\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathbb{N}_0^d, |\boldsymbol{\alpha}| \leq p\}$ , i.e.,

$$d(p) = \sum_{k=0}^p \binom{d+k-1}{k}.$$

$\mathcal{M}_s(\mathcal{X})$  is the set of finite signed measures on  $\mathcal{X}$ .  $\mathcal{L}(\mathcal{X})$  is the set of  $d(p)$ -dimensional finite signed vector measures on  $\mathcal{X}$ . The components of an element of  $\mathcal{L}(\mathcal{X})$  are indexed by multi-indices in  $\mathbb{N}_0^d$  and ordered lexicographically. More specifically, an element  $\Lambda$  of  $\mathcal{L}(\mathcal{X})$  can be denoted by

$$\Lambda = \{\lambda_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathbb{N}_0^d, |\boldsymbol{\alpha}| \leq p\}, \quad (4)$$

where  $\lambda_{\boldsymbol{\alpha}} \in \mathcal{M}_s(\mathcal{X})$  is referred to as the component  $\boldsymbol{\alpha}$  of  $\Lambda$ . The components of  $\Lambda$  follow lexicographical order, i.e.,  $\lambda_{\boldsymbol{\alpha}}$  precedes  $\lambda_{\boldsymbol{\beta}}$  if and only if  $\alpha_i < \beta_i$ ,  $\alpha_j = \beta_j$  for some  $i$  and each  $j$  satisfying  $1 \leq i \leq d$ ,  $1 \leq j < i$ , where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ . For  $\lambda \in \mathcal{M}_s(\mathcal{X})$ ,  $\|\lambda\|$  denotes the total variation norm of  $\lambda$ . For  $\Lambda \in \mathcal{L}(\mathcal{X})$ ,  $\|\Lambda\|$  denotes the total variation norm of  $\Lambda$  induced by the  $l_{\infty}$  vector norm, i.e.,

$$\|\Lambda\| = \max \{\|\lambda_{\boldsymbol{\alpha}}\| : \boldsymbol{\alpha} \in \mathbb{N}_0^d, |\boldsymbol{\alpha}| \leq p\}$$

for  $\Lambda$  specified in (4).  $\mathcal{L}_0(\mathcal{X})$  is the set of  $d(p)$ -dimensional finite vector measures whose component  $\mathbf{0}$  is a probability measure (i.e.,  $\Lambda$  specified in (4) belongs to  $\mathcal{L}_0(\mathcal{X})$  if and only if  $\lambda_{\mathbf{0}} \in \mathcal{P}(\mathcal{X})$ ).

We need a few additional notation:  $r_{\theta,y}^\alpha(x|\lambda)$  and  $s_{\theta,y}^\alpha(x|\Lambda)$  are the functions defined by

$$r_{\theta,y}^\alpha(x|\lambda) = \int \partial_\theta^\alpha r_\theta(y, x|x') \lambda(dx'), \quad s_{\theta,y}^\alpha(x|\Lambda) = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \frac{r_{\theta,y}^{\alpha-\beta}(x|\lambda_\beta)}{\int r_{\theta,y}^0(x'|\lambda_0) \mu(dx')} \quad (5)$$

for  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $\lambda \in \mathcal{M}_s(\mathcal{X})$ ,  $\Lambda = \{\lambda_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq p\} \in \mathcal{L}_0(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .  $f_{\theta,y}^\alpha(x|\Lambda)$  is the function recursively defined by

$$f_{\theta,y}^0(x|\Lambda) = s_{\theta,y}^0(x|\Lambda), \quad f_{\theta,y}^\alpha(x|\Lambda) = s_{\theta,y}^\alpha(x|\Lambda) - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} f_{\theta,y}^\beta(x|\Lambda) \int s_{\theta,y}^{\alpha-\beta}(x'|\Lambda) \mu(dx'), \quad (6)$$

where the recursion is in  $|\alpha|$ .<sup>2</sup>  $R_{\theta,y}^\alpha(dx|\lambda)$ ,  $S_{\theta,y}^\alpha(dx|\Lambda)$  and  $F_{\theta,y}^\alpha(dx|\Lambda)$  are the elements of  $\mathcal{M}_s(\mathcal{X})$  defined by

$$R_{\theta,y}^\alpha(B|\lambda) = \int_B r_{\theta,y}^\alpha(x|\lambda) \mu(dx), \quad S_{\theta,y}^\alpha(B|\Lambda) = \int_B s_{\theta,y}^\alpha(x|\Lambda) \mu(dx), \quad F_{\theta,y}^\alpha(B|\Lambda) = \int_B f_{\theta,y}^\alpha(x|\Lambda) \mu(dx) \quad (7)$$

for  $B \in \mathcal{B}(\mathcal{X})$ , while  $R_{\theta,y}^\alpha(\lambda)$ ,  $S_{\theta,y}^\alpha(\Lambda)$ ,  $F_{\theta,y}^\alpha(\Lambda)$  are a ‘short-hand’ notation for  $R_{\theta,y}^\alpha(dx|\lambda)$ ,  $S_{\theta,y}^\alpha(dx|\Lambda)$ ,  $F_{\theta,y}^\alpha(dx|\Lambda)$  (respectively).  $\langle R_{\theta,y}^\alpha(\lambda) \rangle$ ,  $\langle S_{\theta,y}^\alpha(\Lambda) \rangle$  and  $\langle F_{\theta,y}^\alpha(\Lambda) \rangle$  are the quantities defined by

$$\langle R_{\theta,y}^\alpha(\lambda) \rangle = R_{\theta,y}^\alpha(\mathcal{X}|\lambda), \quad \langle S_{\theta,y}^\alpha(\Lambda) \rangle = S_{\theta,y}^\alpha(\mathcal{X}|\Lambda), \quad \langle F_{\theta,y}^\alpha(\Lambda) \rangle = F_{\theta,y}^\alpha(\mathcal{X}|\Lambda). \quad (8)$$

$F_{\theta,y}(\Lambda)$  is the element of  $\mathcal{L}_0(\mathcal{X})$  defined by

$$F_{\theta,y}(\Lambda) = \{F_{\theta,y}^\alpha(\Lambda) : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}, \quad (9)$$

where  $F_{\theta,y}^\alpha(\Lambda)$  is the component  $\alpha$  of  $F_{\theta,y}(\Lambda)$ .  $F_{\theta,y}^{m:n}(\Lambda)$  is the element of  $\mathcal{L}_0(\mathcal{X})$  recursively defined by

$$F_{\theta,y}^{m:m}(\Lambda) = \Lambda, \quad F_{\theta,y}^{m:n+1}(\Lambda) = F_{\theta,y_{n+1}}(F_{\theta,y}^{m:n}(\Lambda)) \quad (10)$$

for  $n \geq m \geq 0$  and a sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ .  $f_{\theta,y}^{\alpha,m:n}(x|\Lambda)$  is the function defined for  $n > m \geq 0$  by

$$f_{\theta,y}^{\alpha,m:n}(x|\Lambda) = f_{\theta,y_n}^\alpha(x|F_{\theta,y}^{m:n-1}(\Lambda)). \quad (11)$$

$\mathcal{E}_\lambda = \{\mathcal{E}_\lambda^\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$  is the element of  $\mathcal{L}_0(\mathcal{X})$  defined by

$$\mathcal{E}_\lambda^0(B) = \lambda(B), \quad \mathcal{E}_\lambda^\alpha(B) = 0 \quad (12)$$

for  $B \in \mathcal{B}(\mathcal{X})$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $1 \leq |\alpha| \leq p$ , where  $\mathcal{E}_\lambda^0$  and  $\mathcal{E}_\lambda^\alpha$  are (respectively) the component  $\mathbf{0}$  and the component  $\alpha$  of  $\mathcal{E}_\lambda$ .

We will show in Theorem 2.1 that  $F_{\theta,y}^{m:n}(\Lambda)$  is the vector of the optimal filter derivatives of the order up to  $p$ . More specifically, we will demonstrate

$$F_{\theta,y}^{\alpha,m:n}(B|\mathcal{E}_\lambda) = \partial_\theta^\alpha P \left( X_n^{\theta,\lambda} \in B \mid Y_{1:n}^{\theta,\lambda} = y_{1:n} \right)$$

for each  $\theta \in \Theta$ ,  $B \in \mathcal{B}(\mathcal{X})$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ ,  $n \geq 1$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ .

### 2.3. Existence and Stability Results

We analyze here the existence and stability of the optimal filter higher-order derivatives. The analysis is carried out under the following assumptions.

<sup>2</sup>In (6),  $f_{\theta,y}^0(x|\Lambda)$  is the initial condition. At iteration  $k$  of (6) ( $1 \leq k \leq p$ ), function  $f_{\theta,y}^\alpha(x|\Lambda)$  is computed for multi-indices  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = k$  using the results obtained at the previous iterations.

**Assumption 2.1.** *There exists a real number  $\varepsilon \in (0, 1)$  and for each  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ , there exists a measure  $\mu_\theta(dx|y)$  on  $\mathcal{X}$  such that  $0 < \mu_\theta(\mathcal{X}|y) < \infty$  and*

$$\varepsilon \mu_\theta(B|y) \leq \int_B r_\theta(y, x'|x) \mu(dx') \leq \frac{\mu_\theta(B|y)}{\varepsilon}$$

for all  $x \in \mathcal{X}$ ,  $B \in \mathcal{B}(\mathcal{X})$ .

**Assumption 2.2.** *There exists a function  $\psi : \mathcal{Y} \rightarrow [1, \infty)$  such that*

$$|\partial_\theta^\alpha r_\theta(y, x'|x)| \leq (\psi(y))^{|\alpha|} r_\theta(y, x'|x) \quad (13)$$

for all  $\theta \in \Theta$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ ,  $|\alpha| \leq p$ .

**Assumption 2.3.** *There exists a function  $\phi : \mathcal{Y} \times \mathcal{X} \rightarrow [1, \infty)$  such that*

$$r_\theta(y, x'|x) \leq \phi(y, x'), \quad \int \phi(y, x'') \mu(dx'') < \infty$$

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

Assumption 2.1 is a standard strong mixing condition for state-space models. It ensures that the optimal filter  $P_{\theta, \mathbf{y}}^{m:n}(\lambda)$  forgets its initial condition  $\lambda$  exponentially fast (see Proposition 5.2). In this or a similar form, Assumption 2.1 is a crucial ingredient in many results on optimal filtering and statistical inference in state-space and hidden Markov models (see e.g., [1], [2], [5], [6], [8], [10] – [12], [16], [17]; see also [3], [4], [9] and references cited therein). Assumption 2.2 can be considered as an extension of [11, Assumption B] and [18, Assumption 3.2] to the optimal filter higher-order derivatives. It ensures that the higher-order score functions

$$\frac{\partial_\theta^\alpha r_\theta(y, x'|x)}{r_\theta(y, x'|x)}$$

are well-defined and uniformly bounded in  $\theta$ ,  $x, x'$ . Together with Assumptions 2.1 and 2.3, Assumption 2.2 guarantees that the higher-order derivatives of the optimal filter  $P_{\theta, \mathbf{y}}^{m:n}(\lambda)$  exist and forget their initial conditions exponentially fast (see Theorems 2.1 and 2.2). Assumptions 2.1 – 2.3 hold if  $\mathcal{X}$  is a compact set and  $q_\theta(y|x)$  is a mixture of Gaussian densities (see the example studied in Section 4).

Our results on the existence and stability of the optimal filter higher-order derivatives are presented in the next two theorems.

**Theorem 2.1** (Higher-Order Differentiability). *Let Assumptions 2.1 – 2.3 hold. Then,  $p_{\theta, \mathbf{y}}^{m:n}(x|\lambda)$  and  $P_{\theta, \mathbf{y}}^{m:n}(B|\lambda)$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $B \in \mathcal{B}(\mathcal{X})$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $n > m \geq 0$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ . Moreover, we have*

$$\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{m:n}(x|\lambda) = f_{\theta, \mathbf{y}}^{\alpha, m:n}(x|\mathcal{E}_\lambda), \quad \partial_\theta^\alpha P_{\theta, \mathbf{y}}^{m:n}(B|\lambda) = F_{\theta, \mathbf{y}}^{\alpha, m:n}(B|\mathcal{E}_\lambda) \quad (14)$$

for any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .

**Theorem 2.2** (Forgetting). *Let Assumptions 2.1 and 2.2 hold. Then, there exist real numbers  $\tau \in (0, 1)$ ,  $K \in [1, \infty)$  (depending only on  $p, \varepsilon$ ) such that*

$$\|F_{\theta, \mathbf{y}}^{m:n}(\Lambda)\| \leq K \|\Lambda\|^p \left( \sum_{k=m+1}^n \psi(y_k) \right)^p, \quad (15)$$

$$\|F_{\theta, \mathbf{y}}^{m:n}(\Lambda) - F_{\theta, \mathbf{y}}^{m:n}(\Lambda')\| \leq K \tau^{n-m} \|\Lambda - \Lambda'\| (\|\Lambda\| + \|\Lambda'\|)^p \left( \sum_{k=m+1}^n \psi(y_k) \right)^p \quad (16)$$

for all  $\theta \in \Theta$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n \geq m \geq 0$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ .

Theorems 2.1 and 2.2 are proved in Sections 7 and 5, respectively. According to Theorem 2.1, the filtering density  $p_{\theta, \mathbf{y}}^{m:n}(x|\lambda)$  and the filtering distribution  $P_{\theta, \mathbf{y}}^{m:n}(dx|\lambda)$  are  $p$ -times differentiable in  $\theta$ . The same theorem also shows how their higher-order derivatives can be computed recursively using mappings  $f_{\theta, y}^\alpha(x|\Lambda)$ ,  $F_{\theta, y}^\alpha(\Lambda)$ . According to Theorem 2.2, the filtering distribution and its higher-order derivatives  $F_{\theta, \mathbf{y}}^{m:n}(\Lambda)$  forget their initial conditions exponentially fast.

In the rest of the section, we study the ergodicity properties of the optimal filter higher-order derivatives. To do so, we use the following notation.  $\mathcal{Z}$  is the set defined by  $\mathcal{Z} = \mathcal{X} \times \mathcal{Y} \times \mathcal{L}_0(\mathcal{X})$ .  $\Phi_\theta(x, y, \Lambda)$  is a function which maps  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $\Lambda \in \mathcal{L}_0(\mathcal{X})$  to  $\mathbb{R}$ .  $\Phi_\theta(z)$  is another notation for  $\Phi_\theta(x, y, \Lambda)$ , i.e.,  $\Phi_\theta(z) = \Phi_\theta(x, y, \Lambda)$  for  $z = (x, y, \Lambda)$ .  $\{Z_n^{\theta, \Lambda}\}_{n \geq 0}$  and  $\{\tilde{Z}_n^{\theta, \Lambda}\}_{n \geq 0}$  are stochastic processes defined by

$$Z_n^{\theta, \Lambda} = (X_n, Y_n, F_{\theta, \mathbf{Y}}^{0:n}(\Lambda)), \quad \tilde{Z}_n^{\theta, \Lambda} = (X_{n+1}, Y_{n+1}, F_{\theta, \mathbf{Y}}^{0:n}(\Lambda)).$$

for  $n \geq 0$ , where  $\mathbf{Y} = \{Y_n\}_{n \geq 1}$ .  $\Pi_\theta(z, dz')$  and  $\tilde{\Pi}_\theta(z, dz')$  are the kernels on  $\mathcal{Z}$  defined by

$$\begin{aligned} \Pi_\theta(z, B) &= \int \int I_B(x', y', F_{\theta, y'}(\Lambda)) Q(x', dy') P(x, dx'), \\ \tilde{\Pi}_\theta(z, B) &= \int \int I_B(x', y', F_{\theta, y}(\Lambda)) Q(x', dy') P(x, dx') \end{aligned}$$

for  $B \in \mathcal{B}(\mathcal{Z})$  and  $z = (x, y, \Lambda)$ . Then, it is easy to show that  $\{Z_n^{\theta, \Lambda}\}_{n \geq 0}$  and  $\{\tilde{Z}_n^{\theta, \Lambda}\}_{n \geq 0}$  are homogeneous Markov processes whose transition kernels are  $\Pi_\theta(z, dz')$  and  $\tilde{\Pi}_\theta(z, dz')$ , respectively.

To analyze the ergodicity properties of  $\{Z_n^{\theta, \Lambda}\}_{n \geq 0}$  and  $\{\tilde{Z}_n^{\theta, \Lambda}\}_{n \geq 0}$ , we rely on following assumptions.

**Assumption 2.4.** *There exist a probability measure  $\pi(dx)$  on  $\mathcal{X}$  and real numbers  $\delta \in (0, 1)$ ,  $K_0 \in [1, \infty)$  such that*

$$|P^n(x, B) - \pi(B)| \leq K_0 \delta^n$$

for all  $x \in \mathcal{X}$ ,  $B \in \mathcal{B}(\mathcal{X})$ ,  $n \geq 0$ .

**Assumption 2.5.** *There exist a function  $\varphi : \mathcal{X} \times \mathcal{Y} \rightarrow [1, \infty)$  and a real number  $q \in [0, \infty)$  such that*

$$\begin{aligned} |\Phi_\theta(x, y, \Lambda)| &\leq \varphi(x, y) \|\Lambda\|^q, \\ |\Phi_\theta(x, y, \Lambda) - \Phi_\theta(x, y, \Lambda')| &\leq \varphi(x, y) \|\Lambda - \Lambda'\| (\|\Lambda\| + \|\Lambda'\|)^q \end{aligned}$$

for all  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ .

**Assumption 2.6.** *There exists a real number  $L_0 \in [1, \infty)$  such that*

$$\int \varphi(x, y) \psi^r(y) Q(x, dy) \leq L_0 \tag{17}$$

for all  $x \in \mathcal{X}$ , where  $r = p(p + q + 1)$ .

Assumption 2.4 ensures that the Markov process  $\{(X_n, Y_n)\}_{n \geq 0}$  is geometrically ergodic (for further details, see e.g., [14]). Assumption 2.5 is related to function  $\Phi_\theta(x, y, \Lambda)$  and its analytical properties. It requires  $\Phi_\theta(x, y, \Lambda)$  to be locally Lipschitz continuous in  $\Lambda$  and to grow at most polynomially in the same argument. Assumption 2.6 corresponds to the conditional mean of  $\varphi(X_n, Y_n) \psi^r(Y_n)$  given  $X_n = x$ .<sup>3</sup> In this or a similar form, Assumptions 2.4 – 2.6 are involved in many results on the stability of the optimal filter and the asymptotic properties of maximum likelihood estimation in state-space and hidden Markov models (see e.g. [1], [5], [11], [12], [17], [20]; see also [3], [4] and references cited therein).

Our results on the ergodicity of  $\{Z_n^{\theta, \Lambda}\}_{n \geq 0}$  and  $\{\tilde{Z}_n^{\theta, \Lambda}\}_{n \geq 0}$  are presented in the next theorem.

<sup>3</sup>Assumption 2.6 holds under the following conditions: (i)  $\mathcal{X}$  is compact, (ii)  $\varphi(x, y)$  is continuous in  $(x, y)$  and polynomial in  $y$ , (iii)  $\psi(y)$  is polynomial and (iv)  $q_\theta(y|x)$  is Gaussian in  $y$  and continuous in  $(\theta, x, y)$ .

**Theorem 2.3** (Ergodicity). *Let Assumptions 2.1 – 2.6 hold. Moreover, let  $s = p(q + 1)$ . Then, there exist functions  $\phi_\theta, \tilde{\phi}_\theta$  mapping  $\theta \in \Theta$  to  $\mathbb{R}$  such that*

$$\phi_\theta = \lim_{n \rightarrow \infty} (\Pi^n \Phi)_\theta(z), \quad \tilde{\phi}_\theta = \lim_{n \rightarrow \infty} (\tilde{\Pi}^n \Phi)_\theta(z)$$

*for all  $\theta \in \Theta, z \in \mathcal{Z}$ . There also exist real numbers  $\rho \in (0, 1), L \in [1, \infty)$  (depending only on  $\varepsilon, \delta, p, q, K_0, L_0$ ) such that*

$$|(\Pi^n \Phi)_\theta(z) - \phi_\theta| \leq L\rho^n \|\Lambda\|^s, \quad |(\tilde{\Pi}^n \Phi)_\theta(z) - \tilde{\phi}_\theta| \leq L\rho^n \psi^r(y) \|\Lambda\|^s$$

*for all  $\theta \in \Theta, x \in \mathcal{X}, y \in \mathcal{Y}, \Lambda \in \mathcal{L}_0(\mathcal{X}), n \geq 1$  and  $z = (x, y, \Lambda)$ . Here  $(\Pi^n \Phi)_\theta(z)$  and  $(\tilde{\Pi}^n \Phi)_\theta(z)$  are the functions defined by*

$$(\Pi^n \Phi)_\theta(z) = \int \Phi_\theta(z') \Pi_\theta^n(z, dz'), \quad (\tilde{\Pi}^n \Phi)_\theta(z) = \int \Phi_\theta(z') \tilde{\Pi}_\theta^n(z, dz').$$

Theorem 2.3 is proved in Section 6. According to this theorem, Markov processes  $\{Z_n^{\theta, \Lambda}\}_{n \geq 0}$  and  $\{\tilde{Z}_n^{\theta, \Lambda}\}_{n \geq 0}$  are geometrically ergodic. As  $F_{\theta, Y}^{0:n}(\Lambda)$  is a component of  $Z_n^{\theta, \Lambda}$  and  $\tilde{Z}_n^{\theta, \Lambda}$ , the optimal filter and its higher-order derivatives are geometrically ergodic, too.

The optimal filter and its properties have extensively been studied in the literature. However, to the best of our knowledge, the existing results do not provide any information about the existence and stability of the optimal filter higher-order derivatives. Theorems 2.1 – 2.3 fill this gap in the literature on optimal filtering. More specifically, these theorems extend the existing results on the optimal filter first-order derivatives (in particular those of [7], [11] and [18]) to the higher-order derivatives. In Section 3, we use Theorems 2.1 – 2.3 to study the analytical properties of the log-likelihood rate for state-space models. Moreover, in [20], we use the same theorems to analyze the asymptotic behavior of recursive maximum likelihood estimation in state-space models.

### 3. Analytical Properties of Log-Likelihood Rate

In this section, the results presented in Section 2 are used to study the higher-order differentiability of the log-likelihood rate for state-space models. In addition to the notation specified in Section 2, the following notation is used here, too. Let  $q_\theta^n(y_{1:n}|\lambda)$  be the function defined by

$$q_\theta^n(y_{1:n}|\lambda) = \int \cdots \int \int \left( \prod_{k=1}^n r_\theta(y_k, x_k | x_{k-1}) \right) \mu(dx_n) \cdots \mu(dx_1) \lambda(dx_0) \quad (18)$$

for  $\theta \in \Theta, y_1, \dots, y_n \in \mathcal{Y}, \lambda \in \mathcal{P}(\mathcal{X}), n \geq 1$ . Then, the average log-likelihood for state-space model  $\{(X_n, Y_n)\}_{n \geq 0}$  is defined as

$$l_n(\theta, \lambda) = E \left( \frac{1}{n} \log q_\theta^n(Y_{1:n}|\lambda) \right),$$

while the corresponding likelihood rate is the limit  $\lim_{n \rightarrow \infty} l_n(\theta, \lambda)$ . To analyze the analytical and asymptotic properties of  $l_n(\theta, \lambda)$ , we rely on the following assumptions.

**Assumption 3.1.** *There exists a function  $\varphi : \mathcal{Y} \rightarrow [1, \infty)$  such that*

$$|\log \mu_\theta(\mathcal{X}|y)| \leq \varphi(y)$$

*for all  $\theta \in \Theta, y \in \mathcal{Y}$ , where  $\mu_\theta(dx|y)$  is specified in Assumption 2.1.*

**Assumption 3.2.** *There exists a real number  $M_0 \in [1, \infty)$  such that*

$$\int \varphi(y) \psi^u(y) Q(x, dy) \leq M_0, \quad \int \psi^v(y) Q(x, dy) \leq M_0$$

*for all  $x \in \mathcal{X}$ , where  $u = p(p + 1), v = 2p(p + 1)$  and  $\psi(y)$  is specified in Assumption 2.2.*

Assumptions 3.1 and 3.2 are related to the conditional measure  $\mu_\theta(dx|y)$  and its properties. In this or similar form, these assumptions are involved in a number of result on the asymptotic properties of maximum likelihood estimation in state-space and hidden Markov models (see [2], [7], [8], [16], [17]; see also [3] and references cited therein).

Our results on the higher-order differentiability of log-likelihood rate for state-space models are provided in the next theorem.

**Theorem 3.1.** *Let Assumptions 2.1 – 2.4, 3.1 and 3.2 hold. Then, there exists a function  $l : \Theta \rightarrow \mathbb{R}$  which is  $p$ -times differentiable on  $\Theta$  and satisfies  $l(\theta) = \lim_{n \rightarrow \infty} l_n(\theta, \lambda)$  for all  $\theta \in \Theta$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ .*

Theorem 3.1 is proved in Section 7. The theorem claims that the log-likelihood rate  $\lim_{n \rightarrow \infty} l_n(\theta, \lambda)$  is well-defined for each  $\theta \in \Theta$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ . It also claims that this rate is independent of  $\lambda$  and  $p$ -times differentiable in  $\theta$ .

In the context of statistical inference, the properties of log-likelihood rate for state-space and hidden Markov models have been studied in a number of papers (see [2], [7], [8], [16], [17]; see also [3] and references cited therein). However, the existing results do not address the higher-order differentiability of this rate. Theorem 3.1 fills this gap in the literature. Theorem 3.1 is also relevant for asymptotic properties of maximum likelihood estimation in state-space models [20]. The same theorem can also be used to study the higher-order statistical asymptotics for the maximum likelihood estimation in time-series models (for further details on such asymptotics, see e.g. [13], [21]).

## 4. Example

To illustrate the main results, we use them to study optimal filtering in non-linear state-space models. Let  $\Theta$  and  $d$  have the same meaning as in Section 2, while  $\tilde{\Theta} \subseteq \mathbb{R}^d$  is an open set satisfying  $\text{cl}\Theta \subset \tilde{\Theta}$ . We consider the following state-space model:

$$X_{n+1}^{\theta, \lambda} = A_\theta(X_n^{\theta, \lambda}) + B_\theta(X_n^{\theta, \lambda})U_n, \quad Y_n^{\theta, \lambda} = C_\theta(X_n^{\theta, \lambda}) + D_\theta(X_n^{\theta, \lambda})V_n, \quad n \geq 0. \quad (19)$$

Here,  $\theta \in \tilde{\Theta}$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$  are the parameters indexing the model (19).  $A_\theta(x)$  and  $B_\theta(x)$  are functions mapping  $\theta \in \tilde{\Theta}$ ,  $x \in \mathbb{R}^{d_x}$  (respectively) to  $\mathbb{R}^{d_x}$  and  $\mathbb{R}^{d_x \times d_x}$  ( $d_x$  has the same meaning as in Section 2).  $C_\theta(x)$  and  $D_\theta(x)$  are functions mapping  $\theta \in \tilde{\Theta}$ ,  $x \in \mathbb{R}^{d_x}$  (respectively) to  $\mathbb{R}^{d_y}$  and  $\mathbb{R}^{d_y \times d_y}$  ( $d_y$  has the same meaning as in Section 2).  $X_0^{\theta, \lambda}$  is an  $\mathbb{R}^{d_x}$ -valued random variable defined on a probability space  $(\Omega, \mathcal{F}, P)$  and distributed according to  $\lambda$ .  $\{U_n\}_{n \geq 0}$  are  $\mathbb{R}^{d_x}$ -valued i.i.d. random variables which are defined on  $(\Omega, \mathcal{F}, P)$  and have marginal density  $r(u)$  with respect to Lebesgue measure.  $\{V_n\}_{n \geq 0}$  are  $\mathbb{R}^{d_y}$ -valued i.i.d. random variables which are defined on  $(\Omega, \mathcal{F}, P)$  and have marginal density  $s(v)$  with respect to Lebesgue measure. We also assume that  $X_0^{\theta, \lambda}$ ,  $\{U_n\}_{n \geq 0}$  and  $\{V_n\}_{n \geq 0}$  are (jointly) independent.

In addition to the previously introduced notation, the following notation is used here, too.  $\tilde{p}_\theta(x'|x)$  and  $\tilde{q}_\theta(y|x)$  are the functions defined by

$$\tilde{p}_\theta(x'|x) = \frac{r(B_\theta^{-1}(x)(x' - A_\theta(x)))}{|\det B_\theta(x)|}, \quad \tilde{q}_\theta(y|x) = \frac{s(D_\theta^{-1}(x)(y - C_\theta(x)))}{|\det D_\theta(x)|}$$

for  $\theta \in \tilde{\Theta}$ ,  $x, x' \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$  (provided  $B_\theta(x)$  and  $D_\theta(x)$  are invertible).  $p_\theta(x'|x)$  and  $q_\theta(y|x)$  are the functions defined by

$$p_\theta(x'|x) = \frac{r(B_\theta^{-1}(x)(x' - A_\theta(x))) 1_{\mathcal{X}}(x')}{\int_{\mathcal{X}} r(B_\theta^{-1}(x)(x'' - A_\theta(x))) dx''}, \quad q_\theta(y|x) = \frac{s(D_\theta^{-1}(x)(y - C_\theta(x))) 1_{\mathcal{Y}}(y)}{\int_{\mathcal{Y}} s(D_\theta^{-1}(x)(y' - C_\theta(x))) dy'} \quad (20)$$

( $\mathcal{X}$ ,  $\mathcal{Y}$  have the same meaning as in Section 2). It is easy to conclude that  $\tilde{p}_\theta(x'|x)$  and  $\tilde{q}_\theta(y|x)$  are the conditional densities of  $X_{n+1}^{\theta, \lambda}$  and  $Y_n^{\theta, \lambda}$  (respectively) given  $X_n^{\theta, \lambda} = x$ . It is also easy to deduce that  $p_\theta(x'|x)$  and  $q_\theta(y|x)$  accurately approximate  $\tilde{p}_\theta(x'|x)$  and  $\tilde{q}_\theta(y|x)$  when  $\mathcal{X}$  and  $\mathcal{Y}$  are sufficiently large (i.e., when balls of a sufficiently large radius can be inscribed in  $\mathcal{X}$ ,  $\mathcal{Y}$ ).  $p_\theta(x'|x)$  and  $q_\theta(y|x)$  can be interpreted as



truncations of  $\tilde{p}_\theta(x'|x)$  and  $\tilde{q}_\theta(y|x)$  to sets  $\mathcal{X}$  and  $\mathcal{Y}$  (i.e., model specified in (20) can be considered as a truncation of model (19) to  $\mathcal{X}$ ,  $\mathcal{Y}$ ). This or similar truncation is involved (implicitly or explicitly) in the implementation of any numerical approximation to the optimal filter for the model (19).

The optimal filter based on the truncated model (20) is studied under the following assumptions.

**Assumption 4.1.**  $r(x) > 0$  and  $s(y) > 0$  for all  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$ . Moreover,  $B_\theta(x)$  and  $D_\theta(x)$  are invertible for each  $\theta \in \tilde{\Theta}$ ,  $x \in \mathbb{R}^{d_x}$ .

**Assumption 4.2.**  $r(x)$  and  $s(y)$  are  $p$ -times differentiable for all  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$ , where  $p \geq 1$ . Moreover,  $A_\theta(x)$ ,  $B_\theta(x)$ ,  $C_\theta(x)$  and  $D_\theta(x)$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \tilde{\Theta}$ ,  $x \in \mathbb{R}^{d_x}$ .

**Assumption 4.3.**  $\partial^\alpha r(x)$  and  $\partial^\alpha s(y)$  are continuous for each  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ . Moreover,  $\partial_\theta^\alpha A_\theta(x)$ ,  $\partial_\theta^\alpha B_\theta(x)$ ,  $\partial_\theta^\alpha C_\theta(x)$  and  $\partial_\theta^\alpha D_\theta(x)$  are continuous in  $(\theta, x)$  for each  $\theta \in \tilde{\Theta}$ ,  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .

**Assumption 4.4.**  $\mathcal{X}$  and  $\mathcal{Y}$  are compact sets with non-empty interiors.

**Assumption 4.5.**  $\mathcal{X}$  is a compact set with a non-empty interior, while  $\mathcal{Y} = \mathbb{R}^{d_y}$ . Moreover, there exists a real number  $K_0 \in [1, \infty)$  such that

$$s(y) \leq K_0, \quad |\partial^\alpha s(y)| \leq K_0 s(y) (1 + \|y\|)^{|\alpha|}$$

for all  $y \in \mathbb{R}^{d_y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .

**Assumption 4.6.** There exists a real number  $L_0 \in [1, \infty)$  such that

$$|\log s(y)| \leq L_0 (1 + \|y\|)^2$$

for all  $y \in \mathcal{Y}$ .

Assumptions 4.1 – 4.6 cover several classes of non-linear state-space models met in practice — e.g. they hold for a class of stochastic volatility and dynamic probit models. Moreover, these assumptions include non-linear state-space models in which the observation noise  $\{V_n\}_{n \geq 0}$  is a mixture of Gaussian distributions. Other models satisfying assumptions 4.1 – 4.6 can be found in [3], [9] (see also references cited therein).

Our results on the optimal filter for model (20) and its higher-order derivatives read as follows.

**Corollary 4.1.** (i) Let Assumptions 4.1 – 4.4 hold. Then, all conclusions of Theorems 2.1 and 2.2 are true.

(ii) Let Assumptions 2.4, 2.5 and 4.1 – 4.4 hold. Moreover, assume

$$\sup_{x \in \mathcal{X}} \int \varphi(x, y) Q(x, dy) < \infty, \quad (21)$$

where  $\varphi(x, y)$  is specified in Assumption 2.5. Then, all conclusions of Theorem 2.3 are true.

(iii) Let Assumptions 2.4 and 4.1 – 4.4 hold. Then, all conclusions of Theorem 3.1 are true.

**Corollary 4.2.** (i) Let Assumptions 4.1 – 4.3 and 4.5 hold. Then, all conclusions of Theorems 2.1 and 2.2 are true.

(ii) Let Assumptions 2.4, 2.5, 4.1 – 4.3 and 4.5 hold. Moreover, assume

$$\sup_{x \in \mathcal{X}} \int \varphi(x, y) (1 + \|y\|)^{2r} Q(x, dy) < \infty, \quad (22)$$

where  $r$  and  $\varphi(x, y)$  are specified in Assumptions 2.5 and 2.6. Then, all conclusions of Theorem 2.3 are true.

(iii) Let Assumptions 2.4, 4.1 – 4.3, 4.5 and 4.6 hold. Moreover, assume

$$\sup_{x \in \mathcal{X}} \int (1 + \|y\|)^{2v} Q(x, dy) < \infty, \quad (23)$$

where  $v$  is specified in Assumption 3.2. Then, all conclusions of Theorem 3.1 are true.

Corollaries 4.1 and 4.2 are proved in Section 8.

## 5. Proof of Theorem 2.2

In this section, we use the following notation.  $\tau$  is the real number defined as  $\tau = (1 - \varepsilon^2)^{1/2}$ .  $G_{\theta,y}(\lambda, \tilde{\lambda})$  is the element of  $\mathcal{M}_s(\mathcal{X})$  defined by

$$G_{\theta,y}(\lambda, \tilde{\lambda}) = \frac{R_{\theta,y}^0(\tilde{\lambda})}{\langle R_{\theta,y}^0(\lambda) \rangle} - \frac{R_{\theta,y}^0(\lambda) \langle R_{\theta,y}^0(\tilde{\lambda}) \rangle}{\langle R_{\theta,y}^0(\lambda) \rangle^2} \quad (24)$$

for  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $\tilde{\lambda} \in \mathcal{M}_s(\mathcal{X})$ .  $T_{\theta,y}^{\alpha,\beta}(\Lambda)$  is the element of  $\mathcal{M}_s(\mathcal{X})$  defined by

$$T_{\theta,y}^{\alpha,\beta}(\Lambda) = \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - F_{\theta,y}^0(\Lambda) \frac{\langle R_{\theta,y}^{\alpha-\beta}(\lambda_\beta) \rangle}{\langle R_{\theta,y}^0(\lambda_0) \rangle} \quad (25)$$

for  $\Lambda = \{\lambda_\gamma : \gamma \in \mathbb{N}_0^d, |\gamma| \leq p\} \in \mathcal{L}_0(\mathcal{X})$ ,  $\alpha, \beta \in \mathbb{N}_0^d$ ,  $\beta \leq \alpha$ ,  $|\alpha| \leq p$ .  $G_{\theta,y}^\alpha(\Lambda)$  and  $H_{\theta,y}^\alpha(\Lambda)$  are the elements of  $\mathcal{M}_s(\mathcal{X})$  defined by

$$G_{\theta,y}^\alpha(\Lambda) = G_{\theta,y}(\lambda_0, \lambda_\alpha), \quad H_{\theta,y}^\alpha(\Lambda) = \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} T_{\theta,y}^{\alpha,\beta}(\Lambda) - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\mathbf{0}, \alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} F_{\theta,y}^\beta(\Lambda) \langle S_{\theta,y}^{\alpha-\beta}(\Lambda) \rangle. \quad (26)$$

Here and throughout the paper, we rely on the convention that  $\sum_{\beta \in \mathcal{B}}$  is zero whenever  $\mathcal{B} = \emptyset$ . Then, using (5) – (8), it is straightforward to verify

$$S_{\theta,y}^\alpha(\Lambda) = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle}, \quad F_{\theta,y}^\alpha(\Lambda) = S_{\theta,y}^\alpha(\Lambda) - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} F_{\theta,y}^\beta(\Lambda) \langle S_{\theta,y}^{\alpha-\beta}(\Lambda) \rangle. \quad (27)$$

Hence, we get  $F_{\theta,y}^0(\Lambda) = S_{\theta,y}^0(\Lambda) = R_{\theta,y}^0(\lambda_0) / \langle R_{\theta,y}^0(\lambda_0) \rangle$  and

$$T_{\theta,y}^{\alpha,\alpha}(\Lambda) = \frac{R_{\theta,y}^0(\lambda_\alpha)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - F_{\theta,y}^0(\Lambda) \frac{\langle R_{\theta,y}^0(\lambda_\alpha) \rangle}{\langle R_{\theta,y}^0(\lambda_0) \rangle} = \frac{R_{\theta,y}^0(\lambda_\alpha)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - \frac{R_{\theta,y}^0(\lambda_0) \langle R_{\theta,y}^0(\lambda_\alpha) \rangle}{\langle R_{\theta,y}^0(\lambda_0) \rangle^2} = G_{\theta,y}^\alpha(\Lambda).$$

Consequently, (25) – (26) imply

$$\begin{aligned} S_{\theta,y}^\alpha(\Lambda) - F_{\theta,y}^0(\Lambda) \langle S_{\theta,y}^\alpha(\Lambda) \rangle &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \left( \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - F_{\theta,y}^0(\Lambda) \frac{\langle R_{\theta,y}^{\alpha-\beta}(\lambda_\beta) \rangle}{\langle R_{\theta,y}^0(\lambda_0) \rangle} \right) \\ &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} T_{\theta,y}^{\alpha,\beta}(\Lambda) = G_{\theta,y}^\alpha(\Lambda) + \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} T_{\theta,y}^{\alpha,\beta}(\Lambda). \end{aligned} \quad (28)$$

Then, (26), (27) yield

$$F_{\theta,y}^\alpha(\Lambda) = S_{\theta,y}^\alpha(\Lambda) - F_{\theta,y}^0(\Lambda) \langle S_{\theta,y}^\alpha(\Lambda) \rangle - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\mathbf{0}, \alpha\} \\ \beta \leq \alpha}} F_{\theta,y}^\beta(\Lambda) \langle S_{\theta,y}^{\alpha-\beta}(\Lambda) \rangle = G_{\theta,y}^\alpha(\Lambda) + H_{\theta,y}^\alpha(\Lambda). \quad (29)$$

In addition to the previously introduced notation, the following notation is used here, too.  $G_{\theta,y}^{m:n}(\lambda, \tilde{\lambda})$  is the element of  $\mathcal{M}_s(\mathcal{X})$  recursively defined by

$$G_{\theta,y}^{m:m}(\lambda, \tilde{\lambda}) = \tilde{\lambda}, \quad G_{\theta,y}^{m:n}(\lambda, \tilde{\lambda}) = G_{\theta,y_n} \left( P_{\theta,y}^{m:n-1}(\lambda), G_{\theta,y}^{m:n-1}(\lambda, \tilde{\lambda}) \right) \quad (30)$$

for  $\theta \in \Theta$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $\tilde{\lambda} \in \mathcal{M}_s(\mathcal{X})$ ,  $n > m \geq 0$  and a sequence  $y = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ .  $V_{\theta,y}^{\alpha,m:n}(\Lambda)$  and  $W_{\theta,y}^{\alpha,m:n}(\Lambda)$  are the elements of  $\mathcal{M}_s(\mathcal{X})$  defined by

$$V_{\theta,y}^{\alpha,m:n}(\Lambda) = G_{\theta,y}^{m:n}(\lambda_0, \lambda_\alpha), \quad W_{\theta,y}^{\alpha,m:n}(\Lambda) = H_{\theta,y_n}^\alpha(F_{\theta,y}^{m:n-1}(\Lambda)) \quad (31)$$

for  $\Lambda = \{\lambda_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq p\} \in \mathcal{L}_0(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .  $\Phi_{\mathbf{y}}^{m:n}$  and  $\Psi_{\mathbf{y}}^{m:n}$  are the quantities defined by

$$\Phi_{\mathbf{y}}^{m:m} = 1, \quad \Psi_{\mathbf{y}}^{m:m} = 1, \quad \Phi_{\mathbf{y}}^{m:n} = (n-m) \sum_{k=m+1}^n \psi(y_k), \quad \Psi_{\mathbf{y}}^{m:n} = \sum_{k=m+1}^n \psi(y_k).$$

$M_\alpha(\Lambda)$  is the function defined by

$$M_\alpha(\Lambda) = \max \{ \|\lambda_\beta\| : \beta \in \mathbb{N}_0^d, \beta \leq \alpha \}.$$

$K_\alpha(\Lambda, \Lambda')$  and  $L_\alpha(\Lambda, \Lambda')$  are the functions defined by

$$K_\alpha(\Lambda, \Lambda') = \min\{1, M_\alpha(\Lambda - \Lambda')\}, \quad L_\alpha(\Lambda, \Lambda') = M_\alpha(\Lambda) + M_\alpha(\Lambda') \quad (32)$$

for  $\Lambda, \Lambda' \in \mathcal{L}(\mathcal{X})$ .  $L_{\alpha, \mathbf{y}}^{m:n}(\Lambda, \Lambda')$  and  $M_{\alpha, \mathbf{y}}^{m:n}(\Lambda)$  are the functions defined by

$$L_{\alpha, \mathbf{y}}^{m:n}(\Lambda, \Lambda') = (L_\alpha(\Lambda, \Lambda') \Phi_{\mathbf{y}}^{m:n})^{|\alpha|}, \quad M_{\alpha, \mathbf{y}}^{m:n}(\Lambda) = (M_\alpha(\Lambda) \Psi_{\mathbf{y}}^{m:n})^{|\alpha|} \quad (33)$$

for  $n \geq m \geq 0$ .

**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.

**Lemma 5.1.** Let Assumptions 2.1 and 2.2 hold. Then, there exists a real number  $C_1 \in [1, \infty)$  (depending only on  $\varepsilon$ ) such that

$$\left\| \frac{R_{\theta, y}^\alpha(\tilde{\lambda})}{\langle R_{\theta, y}^0(\lambda) \rangle} \right\| \leq C_1 (\psi(y))^{|\alpha|} \|\tilde{\lambda}\|, \quad (34)$$

$$\left\| \frac{R_{\theta, y}^\alpha(\tilde{\lambda})}{\langle R_{\theta, y}^0(\lambda) \rangle} - \frac{R_{\theta, y}^\alpha(\tilde{\lambda}')}{\langle R_{\theta, y}^0(\lambda') \rangle} \right\| \leq C_1 (\psi(y))^{|\alpha|} (\|\tilde{\lambda} - \tilde{\lambda}'\| + \|\lambda - \lambda'\| \|\tilde{\lambda}'\|) \quad (35)$$

for all  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ ,  $\lambda, \lambda' \in \mathcal{P}(\mathcal{X})$ ,  $\tilde{\lambda}, \tilde{\lambda}' \in \mathcal{M}_s(\mathcal{X})$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .

*Proof.* Throughout the proof, we rely on the following notation.  $C_1$  is the real number defined by  $C_1 = \varepsilon^{-4}$  ( $\varepsilon$  is specified in Assumption 2.1).  $\theta, y$  are any elements in  $\Theta, \mathcal{Y}$  (respectively).  $\lambda, \lambda'$  are any elements of  $\mathcal{P}(\mathcal{X})$ , while  $\tilde{\lambda}, \tilde{\lambda}'$  are any elements in  $\mathcal{M}_s(\mathcal{X})$ .  $\alpha$  is any element of  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .

Owing to Assumption 2.1, we have

$$\langle R_{\theta, y}^0(\lambda) \rangle = \int \int r_\theta(y, x' | x) \mu(dx') \lambda(dx) \geq \varepsilon \mu_\theta(\mathcal{X} | y). \quad (36)$$

Moreover, due to Assumptions 2.1, 2.2, we have

$$\begin{aligned} \left\| R_{\theta, y}^\alpha(\tilde{\lambda}) \right\| &\leq \int \int |\partial_\theta^\alpha r_\theta(y, x' | x)| \mu(dx') |\tilde{\lambda}|(dx) \leq (\psi(y))^{|\alpha|} \int \int r_\theta(y, x' | x) \mu(dx') |\tilde{\lambda}|(dx) \\ &\leq \varepsilon^{-1} (\psi(y))^{|\alpha|} \|\tilde{\lambda}\| \mu_\theta(\mathcal{X} | y). \end{aligned} \quad (37)$$

Combining (36), (37), we get

$$\left\| \frac{R_{\theta, y}^\alpha(\tilde{\lambda})}{\langle R_{\theta, y}^0(\lambda) \rangle} \right\| \leq \varepsilon^{-2} (\psi(y))^{|\alpha|} \|\tilde{\lambda}\|. \quad (38)$$

Consequently, we have

$$\begin{aligned} \left\| \frac{R_{\theta, y}^\alpha(\tilde{\lambda})}{\langle R_{\theta, y}^0(\lambda) \rangle} - \frac{R_{\theta, y}^\alpha(\tilde{\lambda}')}{\langle R_{\theta, y}^0(\lambda') \rangle} \right\| &\leq \frac{\left\| R_{\theta, y}^\alpha(\tilde{\lambda}) - R_{\theta, y}^\alpha(\tilde{\lambda}') \right\|}{\langle R_{\theta, y}^0(\lambda) \rangle} + \frac{\left\| R_{\theta, y}^\alpha(\tilde{\lambda}') \right\| \left| \langle R_{\theta, y}^0(\lambda) \rangle - \langle R_{\theta, y}^0(\lambda') \rangle \right|}{\langle R_{\theta, y}^0(\lambda) \rangle \langle R_{\theta, y}^0(\lambda') \rangle} \\ &\leq \frac{\left\| R_{\theta, y}^\alpha(\tilde{\lambda} - \tilde{\lambda}') \right\|}{\langle R_{\theta, y}^0(\lambda') \rangle} + \frac{\left\| R_{\theta, y}^\alpha(\tilde{\lambda}') \right\| \left\| R_{\theta, y}^0(\lambda - \lambda') \right\|}{\langle R_{\theta, y}^0(\lambda) \rangle \langle R_{\theta, y}^0(\lambda') \rangle} \\ &\leq \varepsilon^{-4} (\psi(y))^{|\alpha|} (\|\tilde{\lambda} - \tilde{\lambda}'\| + \|\lambda - \lambda'\| \|\tilde{\lambda}'\|). \end{aligned} \quad (39)$$

Then, (34), (35) directly follow from (38), (39).  $\square$

**Lemma 5.2.** *Let Assumptions 2.1 and 2.2 hold. Then, there exists a real number  $C_2 \in [1, \infty)$  (depending only on  $p, \varepsilon$ ) such that*

$$\|T_{\theta,y}^{\alpha,\beta}(\Lambda)\| \leq C_2 (\psi(y))^{|\alpha-\beta|} \|\lambda_\beta\|, \quad (40)$$

$$\|S_{\theta,y}^\alpha(\Lambda)\| \leq C_2 \sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} (\psi(y))^{|\alpha-\gamma|} \|\lambda_\gamma\|, \quad (41)$$

$$\|T_{\theta,y}^{\alpha,\beta}(\Lambda) - T_{\theta,y}^{\alpha,\beta}(\Lambda')\| \leq C_2 (\psi(y))^{|\alpha-\beta|} (\|\lambda_\beta - \lambda'_\beta\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\beta\|), \quad (42)$$

$$\|S_{\theta,y}^\alpha(\Lambda) - S_{\theta,y}^\alpha(\Lambda')\| \leq C_2 \sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} (\psi(y))^{|\alpha-\gamma|} (\|\lambda_\gamma - \lambda'_\gamma\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\gamma\|) \quad (43)$$

for all  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ ,  $\Lambda = \{\lambda_\gamma : \gamma \in \mathbb{N}_0^d, |\gamma| \leq p\} \in \mathcal{L}_0(\mathcal{X})$ ,  $\Lambda' = \{\lambda'_\gamma : \gamma \in \mathbb{N}_0^d, |\gamma| \leq p\} \in \mathcal{L}_0(\mathcal{X})$  and any multi-indices  $\alpha, \beta \in \mathbb{N}_0^d$ ,  $\beta \leq \alpha$ ,  $|\alpha| \leq p$ .

*Proof.* Throughout the proof, we rely on the following notation.  $C_2$  is the real number defined by  $C_2 = 2^p C_1$  ( $C_1$  is specified in Lemma 5.1).  $\theta, y$  are any elements in  $\Theta, \mathcal{Y}$  (respectively), while  $\Lambda = \{\lambda_\gamma : \gamma \in \mathbb{N}_0^d, |\gamma| \leq p\}$ ,  $\Lambda' = \{\lambda'_\gamma : \gamma \in \mathbb{N}_0^d, |\gamma| \leq p\}$ .  $\alpha, \beta$  are any elements of  $\mathbb{N}_0^d$  satisfying  $\beta \leq \alpha$ ,  $|\alpha| \leq p$ .

Since  $\sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} \binom{\alpha}{\gamma} = 2^{|\alpha|}$ , Lemma 5.1 and (27) imply

$$\|S_{\theta,y}^\alpha(\Lambda)\| \leq \sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} \binom{\alpha}{\gamma} \left\| \frac{R_{\theta,y}^{\alpha-\gamma}(\lambda_\gamma)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} \right\| \leq 2^{|\alpha|} C_1 \sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} (\psi(y))^{|\alpha-\gamma|} \|\lambda_\gamma\|. \quad (44)$$

As  $F_{\theta,y}^0(\Lambda) = R_{\theta,y}^0(\lambda_0) / \langle R_{\theta,y}^0(\lambda_0) \rangle \in \mathcal{P}(\mathcal{X})$ , the same arguments and (25) yield

$$\|T_{\theta,y}^{\alpha,\beta}(\Lambda)\| \leq \left\| \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} \right\| + \|F_{\theta,y}^0(\Lambda)\| \left| \frac{\langle R_{\theta,y}^{\alpha-\beta}(\lambda_\beta) \rangle}{\langle R_{\theta,y}^0(\lambda_0) \rangle} \right| \leq 2 \left\| \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} \right\| \leq 2C_1 (\psi(y))^{|\alpha-\beta|} \|\lambda_\beta\|. \quad (45)$$

Then, (40), (41) directly follow from (44), (45).

Using Lemma 5.1 and (27), we conclude

$$\begin{aligned} \|S_{\theta,y}^\alpha(\Lambda) - S_{\theta,y}^\alpha(\Lambda')\| &\leq \sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} \binom{\alpha}{\gamma} \left\| \frac{R_{\theta,y}^{\alpha-\gamma}(\lambda_\gamma)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - \frac{R_{\theta,y}^{\alpha-\gamma}(\lambda'_\gamma)}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right\| \\ &\leq 2^p C_1 \sum_{\substack{\gamma \in \mathbb{N}_0^d \\ \gamma \leq \alpha}} (\psi(y))^{|\alpha-\gamma|} (\|\lambda_\gamma - \lambda'_\gamma\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\gamma\|). \end{aligned} \quad (46)$$

Relying on the same arguments and (25), we deduce

$$\begin{aligned} \|T_{\theta,y}^{\alpha,\beta}(\Lambda) - T_{\theta,y}^{\alpha,\beta}(\Lambda')\| &\leq \left\| \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - \frac{R_{\theta,y}^{\alpha-\beta}(\lambda'_\beta)}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right\| + \|F_{\theta,y}^0(\Lambda) - F_{\theta,y}^0(\Lambda')\| \left| \frac{\langle R_{\theta,y}^{\alpha-\beta}(\lambda'_\beta) \rangle}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right| \\ &\quad + \|F_{\theta,y}^0(\Lambda)\| \left| \frac{\langle R_{\theta,y}^{\alpha-\beta}(\lambda_\beta) \rangle}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - \frac{\langle R_{\theta,y}^{\alpha-\beta}(\lambda'_\beta) \rangle}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right| \\ &\leq 2 \left\| \frac{R_{\theta,y}^{\alpha-\beta}(\lambda_\beta)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - \frac{R_{\theta,y}^{\alpha-\beta}(\lambda'_\beta)}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right\| + \left\| \frac{R_{\theta,y}^0(\lambda_0)}{\langle R_{\theta,y}^0(\lambda_0) \rangle} - \frac{R_{\theta,y}^0(\lambda'_0)}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right\| \left\| \frac{R_{\theta,y}^{\alpha-\beta}(\lambda'_\beta)}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right\| \\ &\leq 2C_1 (\psi(y))^{|\alpha-\beta|} (\|\lambda_\beta - \lambda'_\beta\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\beta\|). \end{aligned} \quad (47)$$

Then, (42), (43) directly follow from (46), (47).  $\square$

**Proposition 5.1.** *Let Assumption 2.1 hold. Then, there exists a real number  $C_3 \in [1, \infty)$  (depending only on  $\varepsilon$ ) such that*

$$\begin{aligned} \|G_{\theta, \mathbf{y}}^{m:n}(\lambda, \tilde{\lambda})\| &\leq C_3 \tau^{2(n-m)} \|\tilde{\lambda}\|, \\ \|G_{\theta, \mathbf{y}}^{m:n}(\lambda, \tilde{\lambda}) - G_{\theta, \mathbf{y}}^{m:n}(\lambda', \tilde{\lambda}')\| &\leq C_3 \tau^{2(n-m)} (\|\tilde{\lambda} - \tilde{\lambda}'\| + \|\lambda - \lambda'\| \|\tilde{\lambda}'\|) \end{aligned}$$

for all  $\theta \in \Theta$ ,  $\lambda, \lambda' \in \mathcal{P}(\mathcal{X})$ ,  $\tilde{\lambda}, \tilde{\lambda}' \in \mathcal{M}_s(\mathcal{X})$ ,  $n \geq m \geq 0$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$  ( $\tau$  is defined at the beginning of Section 5).

*Proof.* See [18, Lemmas 6.6, 6.7]. □

**Proposition 5.2.** *Let Assumptions 2.1 and 2.2 hold. Then, there exists a real number  $C_4 \in [1, \infty)$  (depending only on  $\varepsilon$ ) such that*

$$\|F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda) - F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda')\| \leq C_4 \tau^{2(n-m)} K_0(\Lambda, \Lambda')$$

for all  $\theta \in \Theta$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n \geq m \geq 0$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$  ( $\tau$  is defined at the beginning of Section 5).

*Proof.* Let  $\theta$  be any element of  $\Theta$ , while  $\mathbf{y} = \{y_n\}_{n \geq 1}$  is any sequence in  $\mathcal{Y}$ . Moreover, let  $\Lambda = \{\lambda_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq p\}$ ,  $\Lambda' = \{\lambda'_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq p\}$  be any elements of  $\mathcal{L}_0(\mathcal{X})$ , while  $n, m$  are any integers satisfying  $n \geq m \geq 0$ .

Using (3), (7), we conclude  $F_{\theta, \mathbf{y}}^{m:m+1}(\lambda_0) = F_{\theta, y_{m+1}}^{\mathbf{0}}(\lambda_0)$ ,  $F_{\theta, \mathbf{y}}^{m:m+1}(\lambda'_0) = F_{\theta, y_{m+1}}^{\mathbf{0}}(\lambda'_0)$  and

$$F_{\theta, \mathbf{y}}^{m:n+1}(\lambda_0) = F_{\theta, y_{n+1}}^{\mathbf{0}}(F_{\theta, \mathbf{y}}^{m:n}(\lambda_0)), \quad F_{\theta, \mathbf{y}}^{m:n+1}(\lambda'_0) = F_{\theta, y_{n+1}}^{\mathbf{0}}(F_{\theta, \mathbf{y}}^{m:n}(\lambda'_0)).$$

Comparing this with (10), we get

$$F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda) = F_{\theta, \mathbf{y}}^{m:n}(\lambda_0), \quad F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda') = F_{\theta, \mathbf{y}}^{m:n}(\lambda'_0)$$

(i.e.,  $F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda)$ ,  $F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda')$  are the filtering distributions initialized by  $\lambda_0, \lambda'_0$ ). Consequently, [18, Theorem 3.1] implies that there exists a real number  $C_4 \in [1, \infty)$  (depending only on  $\varepsilon$ ) such that

$$\|F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda) - F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda')\| \leq C_4 \tau^{2(n-m)} \|\lambda_0 - \lambda'_0\| = C_4 \tau^{2(n-m)} K_0(\Lambda, \Lambda'). \quad (48)$$

□

**Lemma 5.3.** *Let Assumptions 2.1 and 2.2 hold. Then, we have*

$$F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) = V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) + \sum_{k=m+1}^n G_{\theta, \mathbf{y}}^{k:n} \left( F_{\theta, \mathbf{y}}^{\mathbf{0}, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right) \quad (49)$$

for all  $\theta \in \Theta$ ,  $\Lambda \in \mathcal{L}_0(\mathcal{X})$ ,  $n \geq m \geq 0$ , any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ . Here and throughout the paper, we rely on the convention that  $\sum_{k=i}^j$  is zero whenever  $j < i$ .

*Proof.* Throughout the proof, the following notation is used.  $\theta$  is any element of  $\Theta$ , while  $\Lambda = \{\lambda_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq p\}$  is any element of  $\mathcal{L}_0(\mathcal{X})$ .  $m$  is any non-negative integer, while  $\alpha$  is any element of  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .  $\mathbf{y} = \{y_n\}_{n \geq 1}$  is any sequence in  $\mathcal{Y}$ .

We prove (49) by induction in  $n$ . Owing to (10), (30), (31), we have

$$F_{\theta, \mathbf{y}}^{\alpha, m:m}(\Lambda) = \lambda_\alpha, \quad V_{\theta, \mathbf{y}}^{\alpha, m:m}(\Lambda) = G_{\theta, \mathbf{y}}^{m:m}(\lambda_0, \lambda_\alpha) = \lambda_\alpha.$$

Hence, (49) is true when  $n = m$ . Now, suppose that (49) holds for some integer  $n$  satisfying  $n \geq m$ . As  $G_{\theta, \mathbf{y}}(\lambda, \tilde{\lambda})$  is linear in  $\tilde{\lambda}$ , we then get

$$\begin{aligned} G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda), F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) \right) &= G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda), V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) \right) \\ &\quad + \sum_{k=m+1}^n G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{\mathbf{0}, m:n}(\Lambda), G_{\theta, \mathbf{y}}^{k:n} \left( F_{\theta, \mathbf{y}}^{\mathbf{0}, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right) \right). \end{aligned} \quad (50)$$

Since  $P_{\theta, \mathbf{y}}^{m:n}(\lambda_0) = F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda)$  (for further details, see the proof of Proposition 5.2), (30), (31) imply

$$G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda), V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) \right) = G_{\theta, y_{n+1}} \left( P_{\theta, \mathbf{y}}^{m:n}(\lambda_0), G_{\theta, \mathbf{y}}^{m:n}(\lambda_0, \lambda_\alpha) \right) = G_{\theta, \mathbf{y}}^{m:n+1}(\lambda_0, \lambda_\alpha) = V_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda). \quad (51)$$

Moreover, due to (10), we have

$$F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda) = F_{\theta, \mathbf{y}}^{0, k:n}(F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda)) = P_{\theta, \mathbf{y}}^{k:n}(F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda))$$

(for further details, see again the proof of Proposition 5.2). Consequently, (30), (31) yield

$$\begin{aligned} & G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda), G_{\theta, \mathbf{y}}^{k:n} \left( F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right) \right) \\ &= G_{\theta, y_{n+1}} \left( P_{\theta, \mathbf{y}}^{k:n} \left( F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda) \right), G_{\theta, \mathbf{y}}^{k:n} \left( F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right) \right) \\ &= G_{\theta, \mathbf{y}}^{k:n+1} \left( F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right) \end{aligned} \quad (52)$$

for  $n \geq k > m$ . Similarly, (30) implies

$$W_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda) = G_{\theta, \mathbf{y}}^{m+1:n+1} \left( F_{\theta, \mathbf{y}}^{0, m:n+1}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda) \right). \quad (53)$$

Combining (50) – (52), we get

$$G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda), F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) \right) = V_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda) + \sum_{k=m+1}^n G_{\theta, \mathbf{y}}^{k:n+1} \left( F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right).$$

Consequently, (26), (29), (53) imply

$$\begin{aligned} F_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda) &= F_{\theta, y_{n+1}}^{\alpha} \left( F_{\theta, \mathbf{y}}^{m:n}(\Lambda) \right) = G_{\theta, y_{n+1}}^{\alpha} \left( F_{\theta, \mathbf{y}}^{m:n}(\Lambda) \right) + H_{\theta, y_{n+1}}^{\alpha} \left( F_{\theta, \mathbf{y}}^{m:n}(\Lambda) \right) \\ &= G_{\theta, y_{n+1}} \left( F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda), F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) \right) + W_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda) \\ &= V_{\theta, \mathbf{y}}^{\alpha, m:n+1}(\Lambda) + \sum_{k=m+1}^{n+1} G_{\theta, \mathbf{y}}^{k:n+1} \left( F_{\theta, \mathbf{y}}^{0, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda) \right). \end{aligned}$$

Hence, (49) is true for  $n+1$ . Then, the lemma directly follows by the principle of mathematical induction.  $\square$

**Proposition 5.3.** *Let Assumptions 2.1 and 2.2 hold. Then, for each multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ , there exists a real numbers  $A_\alpha \in [1, \infty)$  (depending only on  $p, \varepsilon$ ) such that*

$$\left\| F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) \right\| \leq A_\alpha M_{\alpha, \mathbf{y}}^{m:n}(\Lambda), \quad (54)$$

$$\left\| F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) - F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda') \right\| \leq \tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, \mathbf{y}}^{m:n}(\Lambda, \Lambda') \quad (55)$$

for all  $\theta \in \Theta$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n \geq m \geq 0$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$  ( $\tau$  is defined at the beginning of Section 5).

*Proof.* Throughout the proof, the following notation is used.  $\theta$  is any element of  $\Theta$ , while  $\mathbf{y} = \{y_n\}_{n \geq 1}$  is any sequence in  $\mathcal{Y}$ .  $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3$  are the real numbers defined by

$$\tilde{C}_1 = \frac{4^p C_2 C_3 C_4}{\tau^2 (1 - \tau^2)}, \quad \tilde{C}_2 = \frac{\tilde{C}_1}{4^p}, \quad \tilde{C}_3 = \frac{\tilde{C}_2}{C_3 C_4}$$

( $C_2, C_3, C_4$  are specified in Lemma 5.2 and Propositions 5.1, 5.2).  $A_\alpha$  is the real number defined by  $A_\alpha = \exp(8\tilde{C}_1^2(|\alpha|^2 + 1))$  for  $\alpha \in \mathbb{N}_0^d$ . Then, it easy to show

$$A_\beta \leq \frac{A_\alpha}{\exp(8\tilde{C}_1^2)} \leq \frac{A_\alpha}{8\tilde{C}_1^2}, \quad A_\gamma A_{\alpha-\gamma} \leq \frac{A_\alpha}{\exp(8\tilde{C}_1^2)} \leq \frac{A_\alpha}{8\tilde{C}_1^2} \quad (56)$$

for  $\beta \in \mathbb{N}_0^d \setminus \{\alpha\}$ ,  $\gamma \in \mathbb{N}_0^d \setminus \{0, \alpha\}$ ,  $\beta \leq \alpha$ ,  $\gamma \leq \alpha$ .

Since  $F_{\theta, \mathbf{y}}^{m:n}(\Lambda) = \Lambda$  (due to (10)), (54), (55) are trivially satisfied when  $n = m \geq 0$ . For  $n > m \geq 0$ , we prove (54), (55) by the mathematical induction in  $|\alpha|$ . As  $F_{\theta, \mathbf{y}}^{0, m:n}(\Lambda) \in \mathcal{P}(\mathcal{X})$ , Proposition 5.2 implies that when  $|\alpha| = 0$  (i.e.,  $\alpha = 0$ ), (54), (55) are true for all  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n, m \in \mathbb{N}_0$  fulfilling  $n > m \geq 0$ . Now, the induction hypothesis is formulated: Suppose that (54), (55) hold for some  $l \in \mathbb{N}_0$  and all  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n, m \in \mathbb{N}_0$ ,  $\alpha \in \mathbb{N}_0^d$  satisfying  $0 \leq l < p$ ,  $n > m \geq 0$ ,  $|\alpha| \leq l$ . Then, to prove (54), (55), it is sufficient to show (54), (55) for any  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n, m \in \mathbb{N}_0$ ,  $\alpha \in \mathbb{N}_0^d$  fulfilling  $n > m \geq 0$ ,  $|\alpha| = l + 1$ . In what follows in the proof,  $\Lambda = \{\lambda_\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$ ,  $\Lambda' = \{\lambda'_\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$  are any elements of  $\mathcal{L}_0(\mathcal{X})$ .  $\delta$  is any element of  $\mathbb{N}_0^d$ , while  $\alpha$  is any element of  $\mathbb{N}_0^d$  satisfying  $|\alpha| = l + 1$ .  $\beta, \gamma$  are any elements of  $\mathbb{N}_0^d \setminus \{\alpha\}$  fulfilling  $\beta \leq \alpha$ ,  $\gamma \leq \alpha$ .  $n, m$  are any integers satisfying  $n > m \geq 0$ .

Since  $\beta \leq \alpha$ ,  $\beta \neq \alpha$ , we have  $|\beta| \leq |\alpha| - 1 = l$ . As (54), (55) are trivially satisfied for  $n = m$ , the induction hypothesis imply

$$\max \left\{ \frac{\|F_{\theta, \mathbf{y}}^{\gamma, m:k}(\Lambda)\|}{M_{\gamma, \mathbf{y}}^{m:k}(\Lambda)}, \frac{\|F_{\theta, \mathbf{y}}^{\gamma, m:k}(\Lambda')\|}{M_{\gamma, \mathbf{y}}^{m:k}(\Lambda')} \right\} \leq A_\gamma, \quad (57)$$

$$\frac{\|F_{\theta, \mathbf{y}}^{\gamma, m:k}(\Lambda) - F_{\theta, \mathbf{y}}^{\gamma, m:k}(\Lambda')\|}{L_{\gamma, \mathbf{y}}^{m:k}(\Lambda, \Lambda')} \leq \tau^{2(k-m)} A_\gamma K_\gamma(\Lambda, \Lambda') \quad (58)$$

for  $k \geq m \geq 0$ . Moreover, since  $|\gamma + \delta| = |\gamma| + |\delta|$  and  $M_\gamma(\Lambda) \geq 1$ , (33) yields

$$M_{\gamma, \mathbf{y}}^{m:n-1}(\Lambda) \leq M_{\gamma, \mathbf{y}}^{m:n}(\Lambda) \leq \frac{M_{\gamma+\delta, \mathbf{y}}^{m:n}(\Lambda)}{(\psi(y_n))^{|\delta|}}, \quad M_{\gamma, \mathbf{y}}^{m:n}(\Lambda) M_{\delta, \mathbf{y}}^{m:n}(\Lambda) \leq M_{\gamma+\delta, \mathbf{y}}^{m:n}(\Lambda). \quad (59)$$

Similarly, (33) leads to

$$L_{\gamma, \mathbf{y}}^{m:n-1}(\Lambda, \Lambda') \leq L_{\gamma, \mathbf{y}}^{m:n}(\Lambda, \Lambda') \leq \frac{L_{\gamma+\delta, \mathbf{y}}^{m:n}(\Lambda, \Lambda')}{(\psi(y_n)(n-m))^{|\delta|}}, \quad L_{\gamma, \mathbf{y}}^{m:n}(\Lambda, \Lambda') L_{\delta, \mathbf{y}}^{m:n}(\Lambda, \Lambda') \leq L_{\gamma+\delta, \mathbf{y}}^{m:n}(\Lambda, \Lambda'). \quad (60)$$

The same arguments also imply

$$M_{\delta, \mathbf{y}}^{m:n}(\Lambda) + M_{\delta, \mathbf{y}}^{m:n}(\Lambda') \leq \frac{L_{\delta, \mathbf{y}}^{m:n}(\Lambda, \Lambda')}{(n-m)^{|\delta|}}. \quad (61)$$

Using (57), (59), we conclude

$$\|F_{\theta, \mathbf{y}}^{\gamma, m:n-1}(\Lambda)\| \leq A_\gamma M_{\gamma, \mathbf{y}}^{m:n}(\Lambda) \leq \frac{A_\gamma M_{\gamma+\delta, \mathbf{y}}^{m:n}(\Lambda)}{(\psi(y_n))^{|\delta|}}. \quad (62)$$

Then, Lemma 5.2 and (56), (62) imply

$$\|T_{\theta, y_n}^{\alpha, \beta}(F_{\theta, \mathbf{y}}^{m:n-1}(\Lambda))\| \leq C_2 (\psi(y_n))^{|\alpha-\beta|} \|F_{\theta, \mathbf{y}}^{\beta, m:n-1}(\Lambda)\| \leq C_2 A_\beta M_{\alpha, \mathbf{y}}^{m:n}(\Lambda) \leq \frac{A_\alpha M_{\alpha, \mathbf{y}}^{m:n}(\Lambda)}{2\tilde{C}_1} \quad (63)$$

(as  $C_2 \leq 4\tilde{C}_1$ ). The same lemma and (62) yield

$$\|S_{\theta, y_n}^\gamma(F_{\theta, \mathbf{y}}^{m:n-1}(\Lambda))\| \leq C_2 \sum_{\substack{\delta \in \mathbb{N}_0^d \\ \delta \leq \gamma}} (\psi(y_n))^{|\gamma-\delta|} \|F_{\theta, \mathbf{y}}^{\delta, m:n-1}(\Lambda)\| \leq 2^{|\gamma|} C_2 A_\gamma M_{\gamma, \mathbf{y}}^{m:n}(\Lambda) \leq \tilde{C}_1 A_\gamma M_{\gamma, \mathbf{y}}^{m:n}(\Lambda) \quad (64)$$

(since  $2^{|\gamma|} C_2 \leq 2^p C_2 \leq \tilde{C}_1/2$ ). If  $\beta \neq 0$ , (56), (57), (59), (64) lead to

$$\begin{aligned} \|F_{\theta, \mathbf{y}}^{\beta, m:n}(\Lambda) \langle S_{\theta, y_n}^{\alpha-\beta}(F_{\theta, \mathbf{y}}^{m:n-1}(\Lambda)) \rangle\| &\leq \|F_{\theta, \mathbf{y}}^{\beta, m:n}(\Lambda)\| \|S_{\theta, y_n}^{\alpha-\beta}(F_{\theta, \mathbf{y}}^{m:n-1}(\Lambda))\| \\ &\leq \tilde{C}_1 A_\beta A_{\alpha-\beta} M_{\beta, \mathbf{y}}^{m:n}(\Lambda) M_{\alpha-\beta, \mathbf{y}}^{m:n}(\Lambda) \\ &\leq \frac{A_\alpha M_{\alpha, \mathbf{y}}^{m:n}(\Lambda)}{2\tilde{C}_1}. \end{aligned} \quad (65)$$

Consequently, (10), (26) (31), (63) imply

$$\begin{aligned}
\|W_{\theta, \mathbf{y}}^{\alpha, m: n}(\Lambda)\| &\leq \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \|T_{\theta, \mathbf{y}_n}^{\alpha, \beta}(F_{\theta, \mathbf{y}}^{m: n-1}(\Lambda))\| + \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\mathbf{0}, \alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \|F_{\theta, \mathbf{y}}^{\beta, m: n}(\Lambda) \langle S_{\theta, \mathbf{y}_n}^{\alpha - \beta}(F_{\theta, \mathbf{y}}^{m: n-1}(\Lambda)) \rangle\| \\
&\leq \frac{2^{|\alpha|} A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)}{\tilde{C}_1} \\
&\leq \frac{A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)}{\tilde{C}_2}
\end{aligned} \tag{66}$$

(as  $\tilde{C}_1/2^{|\alpha|} \geq \tilde{C}_1/2^p \geq \tilde{C}_2$ ). Then, owing to Proposition 5.1, we have

$$\begin{aligned}
\|G_{\theta, \mathbf{y}}^{k: n}(F_{\theta, \mathbf{y}}^{\mathbf{0}, m: k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m: k}(\Lambda))\| &\leq C_3 \tau^{2(n-k)} \|W_{\theta, \mathbf{y}}^{\alpha, m: k}(\Lambda)\| \leq \frac{C_3 \tau^{2(n-k)} A_{\alpha} M_{\alpha, \mathbf{y}}^{m: k}(\Lambda)}{\tilde{C}_2} \\
&\leq \frac{\tau^{2(n-k)} A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)}{\tilde{C}_3}
\end{aligned} \tag{67}$$

for  $n \geq k > m$  (since  $C_3/\tilde{C}_2 \leq 1/\tilde{C}_3$ ). Due to the same proposition and (31), we have

$$\|V_{\theta, \mathbf{y}}^{\alpha, m: n}(\Lambda)\| \leq C_3 \tau^{2(n-m)} \|\lambda_{\alpha}\| \leq C_3 \tau^{2(n-m)} M_{\alpha}(\Lambda) \leq \frac{\tau^{2(n-m)} A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)}{\tilde{C}_3} \tag{68}$$

(as  $A_{\alpha} \geq \tilde{C}_1^2 \geq C_3 \tilde{C}_3$ ). Combining Lemma 5.3 and (67), (68), we get

$$\begin{aligned}
\|F_{\theta, \mathbf{y}}^{\alpha, m: n}(\Lambda)\| &\leq \|V_{\theta, \mathbf{y}}^{m: n}(\Lambda)\| + \sum_{k=m+1}^n \|G_{\theta, \mathbf{y}}^{k: n}(F_{\theta, \mathbf{y}}^{\mathbf{0}, m: k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m: k}(\Lambda))\| \leq \frac{A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)}{\tilde{C}_3} \sum_{k=m}^n \tau^{2(n-k)} \\
&\leq \frac{A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)}{\tilde{C}_3(1 - \tau^2)} \\
&\leq A_{\alpha} M_{\alpha, \mathbf{y}}^{m: n}(\Lambda)
\end{aligned} \tag{69}$$

(since  $\tilde{C}_3(1 - \tau^2) \geq 1$ ). Hence, (54) holds for  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = l + 1$ .

Now, (55) is proved. Relying on (58), (60), we deduce

$$\begin{aligned}
\|F_{\theta, \mathbf{y}}^{\gamma, m: n-1}(\Lambda) - F_{\theta, \mathbf{y}}^{\gamma, m: n-1}(\Lambda')\| &\leq \tau^{2(n-m-1)} A_{\gamma} K_{\gamma}(\Lambda, \Lambda') L_{\gamma, \mathbf{y}}^{m: n}(\Lambda, \Lambda') \\
&\leq \frac{\tau^{2(n-m-1)} A_{\gamma} K_{\gamma + \delta}(\Lambda, \Lambda') L_{\gamma + \delta, \mathbf{y}}^{m: n}(\Lambda, \Lambda')}{(\psi(y_n)(n-m))^{|\delta|}}.
\end{aligned} \tag{70}$$

Similarly, using Proposition 5.2 and (60), (62), we conclude

$$\begin{aligned}
\|F_{\theta, \mathbf{y}}^{\mathbf{0}, m: n-1}(\Lambda) - F_{\theta, \mathbf{y}}^{\mathbf{0}, m: n-1}(\Lambda')\| &\leq \|F_{\theta, \mathbf{y}}^{\gamma, m: n-1}(\Lambda')\| \leq C_4 \tau^{2(n-m-1)} A_{\gamma} K_{\mathbf{0}}(\Lambda, \Lambda') M_{\gamma, \mathbf{y}}^{m: n}(\Lambda') \\
&\leq \frac{C_4 \tau^{2(n-m-1)} A_{\gamma} K_{\gamma + \delta}(\Lambda, \Lambda') L_{\gamma + \delta, \mathbf{y}}^{m: n}(\Lambda, \Lambda')}{(\psi(y_n)(n-m))^{|\delta|}}
\end{aligned} \tag{71}$$

(since  $M_{\gamma, \mathbf{y}}^{m: n}(\Lambda') \leq L_{\gamma, \mathbf{y}}^{m: n}(\Lambda, \Lambda')$ ). Then, Lemma 5.2 and (56) imply

$$\begin{aligned}
&\|T_{\theta, \mathbf{y}_n}^{\alpha, \beta}(F_{\theta, \mathbf{y}}^{m: n-1}(\Lambda)) - T_{\theta, \mathbf{y}_n}^{\alpha, \beta}(F_{\theta, \mathbf{y}}^{m: n-1}(\Lambda'))\| \leq C_2 (\psi(y_n))^{|\alpha - \beta|} \|F_{\theta, \mathbf{y}}^{\beta, m: n-1}(\Lambda) - F_{\theta, \mathbf{y}}^{\beta, m: n-1}(\Lambda')\| \\
&\quad + C_2 (\psi(y_n))^{|\alpha - \beta|} \|F_{\theta, \mathbf{y}}^{\mathbf{0}, m: n-1}(\Lambda) - F_{\theta, \mathbf{y}}^{\mathbf{0}, m: n-1}(\Lambda')\| \|F_{\theta, \mathbf{y}}^{\beta, m: n-1}(\Lambda')\| \\
&\leq \frac{2C_2 C_4 \tau^{2(n-m-1)} A_{\beta} K_{\alpha}(\Lambda, \Lambda') L_{\alpha, \mathbf{y}}^{m: n}(\Lambda, \Lambda')}{(n-m)^{|\alpha - \beta|}} \\
&\leq \frac{\tau^{2(n-m)} A_{\alpha} K_{\alpha}(\Lambda, \Lambda') L_{\alpha, \mathbf{y}}^{m: n}(\Lambda, \Lambda')}{4\tilde{C}_1(n-m)}
\end{aligned} \tag{72}$$



(as  $|\alpha - \beta| \geq 1$ ,  $C_2 C_4 \leq \tilde{C}_1 \tau^2$ ). The same lemma and (70), (71) yield

$$\begin{aligned}
& \left\| S_{\theta, y_n}^\gamma (F_{\theta, y}^{m:n-1}(\Lambda)) - S_{\theta, y_n}^\gamma (F_{\theta, y}^{m:n-1}(\Lambda')) \right\| \leq C_2 \sum_{\substack{\delta \in \mathbb{N}_0^d \\ \delta \leq \gamma}} (\psi(y_n))^{|\gamma - \delta|} \left\| F_{\theta, y}^{\delta, m:n-1}(\Lambda) - F_{\theta, y}^{\delta, m:n-1}(\Lambda') \right\| \\
& + C_2 \sum_{\substack{\delta \in \mathbb{N}_0^d \\ \delta \leq \gamma}} (\psi(y_n))^{|\gamma - \delta|} \left\| F_{\theta, y}^{\mathbf{0}, m:n-1}(\Lambda) - F_{\theta, y}^{\mathbf{0}, m:n-1}(\Lambda') \right\| \left\| F_{\theta, y}^{\delta, m:n-1}(\Lambda') \right\| \\
& \leq 4^{|\gamma|} C_2 C_4 \tau^{2(n-m-1)} A_\gamma K_\gamma(\Lambda, \Lambda') L_{\gamma, y}^{m:n}(\Lambda, \Lambda') \\
& \leq \tilde{C}_1 \tau^{2(n-m)} A_\gamma K_\gamma(\Lambda, \Lambda') L_{\gamma, y}^{m:n}(\Lambda, \Lambda')
\end{aligned} \tag{73}$$

(since  $4^{|\gamma|} C_2 C_4 \leq \tilde{C}_1 \tau^2$ ). Then, (57), (58), (64) lead to

$$\begin{aligned}
& \left\| F_{\theta, y}^{\beta, m:n}(\Lambda) \langle S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda)) \rangle - F_{\theta, y}^{\beta, m:n}(\Lambda') \langle S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda')) \rangle \right\| \\
& \leq \left\| F_{\theta, y}^{\beta, m:n}(\Lambda) \right\| \left\| S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda)) - S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda')) \right\| \\
& \quad + \left\| F_{\theta, y}^{\beta, m:n}(\Lambda) - F_{\theta, y}^{\beta, m:n}(\Lambda') \right\| \left\| S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda')) \right\| \\
& \leq \tilde{C}_1 \tau^{2(n-m)} A_\beta A_{\alpha - \beta} K_{\alpha - \beta}(\Lambda, \Lambda') L_{\alpha - \beta, y}^{m:n}(\Lambda, \Lambda') M_{\beta, y}^{m:n}(\Lambda) \\
& \quad + \tilde{C}_1 \tau^{2(n-m)} A_\beta A_{\alpha - \beta} K_\beta(\Lambda, \Lambda') L_{\beta, y}^{m:n}(\Lambda, \Lambda') M_{\alpha - \beta, y}^{m:n}(\Lambda').
\end{aligned}$$

Hence, if  $\beta \neq \mathbf{0}$ , (56), (60), (61) imply

$$\begin{aligned}
& \left\| F_{\theta, y}^{\beta, m:n}(\Lambda) \langle S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda)) \rangle - F_{\theta, y}^{\beta, m:n}(\Lambda') \langle S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda')) \rangle \right\| \\
& \leq \frac{\tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, y}^{m:n}(\Lambda, \Lambda')}{4\tilde{C}_1(n-m)}
\end{aligned} \tag{74}$$

(as  $\beta \neq \mathbf{0}$ ,  $\alpha - \beta \neq \mathbf{0}$ ). Consequently, (10), (26), (31), (72) yield

$$\begin{aligned}
& \left\| W_{\theta, y}^{\alpha, m:n}(\Lambda) - W_{\theta, y}^{\alpha, m:n}(\Lambda') \right\| \leq \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \left\| T_{\theta, y_n}^{\alpha, \beta}(F_{\theta, y}^{m:n-1}(\Lambda)) - T_{\theta, y_n}^{\alpha, \beta}(F_{\theta, y}^{m:n-1}(\Lambda')) \right\| \\
& + \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\mathbf{0}, \alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \left\| F_{\theta, y}^{\beta, m:n}(\Lambda) \langle S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda)) \rangle - F_{\theta, y}^{\beta, m:n}(\Lambda') \langle S_{\theta, y_n}^{\alpha - \beta}(F_{\theta, y}^{m:n-1}(\Lambda')) \rangle \right\| \\
& \leq \frac{4^{|\alpha|} \tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, y}^{m:n}(\Lambda, \Lambda')}{2\tilde{C}_1(n-m)} \\
& \leq \frac{\tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, y}^{m:n}(\Lambda, \Lambda')}{2\tilde{C}_2(n-m)}
\end{aligned} \tag{75}$$

(since  $\tilde{C}_1/4^{|\alpha|} \geq \tilde{C}_1/4^p = \tilde{C}_2$ ). Then, owing to Propositions 5.1, 5.2 and (61), (66), we have

$$\begin{aligned}
& \left\| G_{\theta, y}^{k:n} \left( F_{\theta, y}^{\mathbf{0}, m:k}(\Lambda), W_{\theta, y}^{\alpha, m:k}(\Lambda) \right) - G_{\theta, y}^{k:n} \left( F_{\theta, y}^{\mathbf{0}, m:k}(\Lambda'), W_{\theta, y}^{\alpha, m:k}(\Lambda') \right) \right\| \\
& \leq C_3 \tau^{2(n-k)} \left\| W_{\theta, y}^{\alpha, m:k}(\Lambda) - W_{\theta, y}^{\alpha, m:k}(\Lambda') \right\| + C_3 \tau^{2(n-k)} \left\| F_{\theta, y}^{\mathbf{0}, m:k}(\Lambda) - F_{\theta, y}^{\mathbf{0}, m:k}(\Lambda') \right\| \left\| W_{\theta, y}^{\alpha, m:k}(\Lambda') \right\| \\
& \leq \frac{C_3 \tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, y}^{m:k}(\Lambda, \Lambda')}{2\tilde{C}_2(n-m)} + \frac{C_3 C_4 \tau^{2(n-m)} A_\alpha K_{\mathbf{0}}(\Lambda, \Lambda') M_{\alpha, y}^{m:k}(\Lambda')}{2\tilde{C}_2} \\
& \leq \frac{\tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, y}^{m:n}(\Lambda, \Lambda')}{\tilde{C}_3(n-m)}
\end{aligned} \tag{76}$$

for  $n \geq k > m$  (as  $C_3/\tilde{C}_2 \leq C_3 C_4/\tilde{C}_2 = 1/\tilde{C}_3$ ). Due to the same propositions and (31), we have

$$\left\| V_{\theta, y}^{\alpha, m:n}(\Lambda) - V_{\theta, y}^{\alpha, m:n}(\Lambda') \right\| \leq C_3 \tau^{2(n-m)} (\|\lambda_\alpha - \lambda'_\alpha\| + \|\lambda_{\mathbf{0}} - \lambda'_{\mathbf{0}}\| \|\lambda'_\alpha\|). \tag{77}$$

Moreover, we have

$$\begin{aligned}\|\lambda_\alpha - \lambda'_\alpha\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\alpha\| &\leq M_\alpha(\Lambda - \Lambda') + M_\alpha(\Lambda - \Lambda') M_\alpha(\Lambda') \leq 2M_\alpha(\Lambda - \Lambda') L_\alpha(\Lambda, \Lambda'), \\ \|\lambda_\alpha - \lambda'_\alpha\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\alpha\| &\leq \|\lambda_\alpha\| + 2\|\lambda'_\alpha\| \leq 3L_\alpha(\Lambda, \Lambda')\end{aligned}$$

(since  $M_\alpha(\Lambda') \geq \|\lambda'_0\| = 1$ ). Hence, we get

$$\|\lambda_\alpha - \lambda'_\alpha\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\alpha\| \leq 3 \min\{1, M_\alpha(\Lambda - \Lambda')\} L_\alpha(\Lambda, \Lambda') = 3K_\alpha(\Lambda, \Lambda') L_\alpha(\Lambda, \Lambda').$$

Therefore, (77) implies

$$\begin{aligned}\|V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) - V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda')\| &\leq 3C_3 \tau^{2(n-m)} K_\alpha(\Lambda, \Lambda') L_\alpha(\Lambda, \Lambda') \\ &\leq \frac{\tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, \mathbf{y}}^{m:n}(\Lambda, \Lambda')}{\tilde{C}_3}\end{aligned}\quad (78)$$

(as  $A_\alpha \geq \tilde{C}_1^2 \geq 3C_3 \tilde{C}_3$ ). Combining Lemma 5.3 and (76), (78), we get

$$\begin{aligned}\|F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) - F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda')\| &\leq \|V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) - V_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda')\| \\ &\quad + \sum_{k=m+1}^n \|G_{\theta, \mathbf{y}}^{k:n}(F_{\theta, \mathbf{y}}^{\mathbf{0}, m:k}(\Lambda), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda)) - G_{\theta, \mathbf{y}}^{k:n}(F_{\theta, \mathbf{y}}^{\mathbf{0}, m:k}(\Lambda'), W_{\theta, \mathbf{y}}^{\alpha, m:k}(\Lambda'))\| \\ &\leq \frac{2\tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, \mathbf{y}}^{m:n}(\Lambda, \Lambda')}{\tilde{C}_3} \\ &\leq \tau^{2(n-m)} A_\alpha K_\alpha(\Lambda, \Lambda') L_{\alpha, \mathbf{y}}^{m:n}(\Lambda, \Lambda')\end{aligned}\quad (79)$$

(since  $\tilde{C}_3 \geq 2$ ). Hence, (55) holds for  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = l + 1$ . Then, the proposition directly follows by the principle of mathematical induction.  $\square$

**Proof of Theorem 2.2.** Let  $\tilde{C}_1, \tilde{C}_2$  be the real numbers defined by  $\tilde{C}_1 = \max_{n \geq 1} \tau^{n-1} n^p$ ,  $\tilde{C}_2 = \max\{A_\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$ , while  $K$  is the real number defined by  $K = \tilde{C}_1 \tilde{C}_2$  ( $A_\alpha$  is specified in Proposition 5.3, while  $\tau$  is defined at the beginning of Section 5). Then, Proposition 5.3 implies

$$\begin{aligned}\|F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda) - F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda')\| &\leq \tilde{C}_2 \tau^{2(n-m)} (n-m)^p \|\Lambda - \Lambda'\| ((\|\Lambda\| + \|\Lambda'\|) \Psi_{\mathbf{y}}^{m:n})^p \\ &\leq \tilde{C}_1 \tilde{C}_2 \tau^{n-m} \|\Lambda - \Lambda'\| ((\|\Lambda\| + \|\Lambda'\|) \Psi_{\mathbf{y}}^{m:n})^p \\ &\leq K \tau^{n-m} \|\Lambda - \Lambda'\| (\|\Lambda\| + \|\Lambda'\|)^p \left( \sum_{k=m+1}^n \psi(y_k) \right)^p\end{aligned}\quad (80)$$

for  $\theta \in \Theta$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $n > m \geq 0$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$  and a sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$ . Proposition 5.3 also yields

$$\|F_{\theta, \mathbf{y}}^{\alpha, m:n}(\Lambda)\| \leq \tilde{C}_2 (\|\Lambda\| \Psi_{\mathbf{y}}^{m:n})^p \leq K \|\Lambda\|^p \left( \sum_{k=m+1}^n \psi(y_k) \right)^p\quad (81)$$

for the same  $\theta, \Lambda, n, m, \alpha, \mathbf{y}$ . As (15), (16) are trivially satisfied when  $n = m$ , the theorem directly follows from (80), (81).  $\square$

## 6. Proof of Theorem 2.3

In this section, we rely on the following notation.  $\tilde{\Phi}_\theta(x, y, \Lambda)$  is the function defined by

$$\tilde{\Phi}_\theta(x, y, \Lambda) = \int \int \Phi_\theta(x', y', \Lambda) Q(x', dy') P(x, dx')\quad (82)$$

for  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $\Lambda \in \mathcal{L}_0(\mathcal{X})$ .  $\mathbf{X}$  and  $\mathbf{Y}$  denote stochastic processes  $\{X_n\}_{n \geq 1}$  and  $\{Y_n\}_{n \geq 1}$  (i.e.,  $\mathbf{X} = \{X_n\}_{n \geq 1}$ ,  $\mathbf{Y} = \{Y_n\}_{n \geq 1}$ ).  $G_{\theta, \mathbf{X}, \mathbf{Y}}^{m:n}(\Lambda)$  and  $H_{\theta, \mathbf{X}, \mathbf{Y}}^{m:n}(\Lambda)$  are the random functions defined by

$$G_{\theta, \mathbf{X}, \mathbf{Y}}^{m:n}(\Lambda) = \Phi_\theta(X_n, Y_n, F_{\theta, \mathbf{Y}}^{m:n}(\Lambda)), \quad H_{\theta, \mathbf{X}, \mathbf{Y}}^{m:n}(\Lambda) = \Phi_\theta(X_{n+1}, Y_{n+1}, F_{\theta, \mathbf{Y}}^{m:n}(\Lambda))$$

for  $n \geq m \geq 0$ .  $A_\theta^n(x, \Lambda)$  and  $B_\theta^n(x, \Lambda)$  are the functions defined by

$$A_\theta^n(x, \Lambda) = E(G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda) | X_0 = x), \quad B_\theta^n(x, \Lambda) = E(G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) | X_0 = x)$$

for  $n \geq 1$ .  $C_\theta^n(x, y, \Lambda)$  and  $D_\theta^n(x, y, \Lambda)$  are the functions defined by

$$C_\theta^n(x, y, \Lambda) = E(H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda) | X_1 = x, Y_1 = y), \\ D_\theta^n(x, y, \Lambda) = E(H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) | X_1 = x, Y_1 = y).$$

$\tilde{A}_\theta^{m,n}(x, \Lambda)$  and  $\tilde{B}_\theta^n(x, \Lambda)$  are the functions defined by

$$\tilde{A}_\theta^{m,n}(x, \Lambda) = \int A_\theta^{n-m}(x', \Lambda)(P^m - \pi)(x, dx'), \\ \tilde{B}_\theta^n(x, \Lambda) = \int \int \Phi_\theta(x', y', \Lambda)Q(x', dy')(P^n - \pi)(x, dx')$$

for  $n > m \geq 0$ .

**Lemma 6.1.** *Let Assumptions 2.1, 2.2 and 2.5 hold. Then, there exists a real number  $C_5 \in [1, \infty)$  (depending only on  $p, q, \varepsilon, L_0$ ) such that*

$$\max\{|G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda)|, |G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda)|\} \leq C_5 \tau^n \|\Lambda\|^s \varphi(X_n, Y_n) \sum_{k=1}^n \psi^r(Y_k), \\ \max\{|H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda)|, |H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda)|\} \leq C_5 \tau^n \|\Lambda\|^s \varphi(X_{n+1}, Y_{n+1}) \sum_{k=1}^n \psi^r(Y_k)$$

for all  $\theta \in \Theta$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $\Lambda \in \mathcal{L}_0(\mathcal{X})$ ,  $n \geq 1$  ( $r$  and  $s$  are specified in Assumption 2.6 and Theorem 2.3, while  $\tau$  is defined at the beginning of Section 5).

*Proof.* Throughout the proof, the following notation is used.  $\tilde{C}_1, \tilde{C}_2$  are the real numbers defined by  $\tilde{C}_1 = \max_{n \geq 1} \tau^{n-1} n^{2r}$ ,  $\tilde{C}_2 = \max\{A_\alpha : \alpha \in \mathbb{N}_0, |\alpha| \leq p\}$  ( $A_\alpha$  is specified in Proposition 5.3).  $\tilde{C}_3, \tilde{C}_4$  are the real numbers defined by  $\tilde{C}_3 = 2^p \tilde{C}_2^{p+1}$ ,  $\tilde{C}_4 = 2^q \tilde{C}_2^q \tilde{C}_3$ , while  $C_5$  is the real number defined by  $C_5 = \tilde{C}_1 \tilde{C}_4 \tau^{-2}$ .  $\theta, x, y, \lambda$  are any elements of  $\Theta, \mathcal{X}, \mathcal{Y}, \mathcal{P}(\mathcal{X})$  (respectively), while  $\Lambda, \Lambda'$  are any elements of  $\mathcal{L}_0(\mathcal{X})$ .  $\mathbf{y} = \{y_n\}_{n \geq 1}$  is any sequence in  $\mathcal{Y}$ .  $n, m, k$  are any integers satisfying  $n \geq 1, k \geq m \geq 0$ .

Owing to Proposition 5.3, we have

$$\|F_{\theta, \mathbf{y}}^{m:k}(\Lambda)\| \leq \tilde{C}_2 (\|\Lambda\| \Phi_{\mathbf{y}}^{0:k})^p, \quad \|F_{\theta, \mathbf{y}}^{m:k}(\Lambda) - F_{\theta, \mathbf{y}}^{m:k}(\Lambda')\| \leq \tilde{C}_2 \tau^{2(k-m)} ((\|\Lambda\| + \|\Lambda'\|) \Phi_{\mathbf{y}}^{0:k})^p \quad (83)$$

(as  $\Phi_{\mathbf{y}}^{0:k} \geq \Phi_{\mathbf{y}}^{m:k} \geq \Psi_{\mathbf{y}}^{m:k}$ ). Consequently, we have

$$\|F_{\theta, \mathbf{y}}^{0:k}(\mathcal{E}_\lambda)\| + \|F_{\theta, \mathbf{y}}^{m:k}(\Lambda)\| \leq 2\tilde{C}_2 (\|\Lambda\| \Phi_{\mathbf{y}}^{0:k})^p, \\ \|F_{\theta, \mathbf{y}}^{0:m}(\mathcal{E}_\lambda)\| + \|\Lambda\| \leq 2\tilde{C}_2 (\|\Lambda\| \Phi_{\mathbf{y}}^{0:m})^p \leq 2\tilde{C}_2 (\|\Lambda\| \Phi_{\mathbf{y}}^{0:k})^p. \quad (84)$$

Then, (10), (83) imply

$$\|F_{\theta, \mathbf{y}}^{0:k}(\mathcal{E}_\lambda) - F_{\theta, \mathbf{y}}^{m:k}(\Lambda)\| = \|F_{\theta, \mathbf{y}}^{m:k}(F_{\theta, \mathbf{y}}^{0:m}(\mathcal{E}_\lambda)) - F_{\theta, \mathbf{y}}^{m:k}(\Lambda)\| \leq \tilde{C}_2 \tau^{2(k-m)} ((\|F_{\theta, \mathbf{y}}^{0:m}(\mathcal{E}_\lambda)\| + \|\Lambda\|) \Phi_{\mathbf{y}}^{0:k})^p \\ \leq \tilde{C}_3 \tau^{2(k-m)} \|\Lambda\|^{p^2} (\Phi_{\mathbf{y}}^{0:k})^{p(p+1)} \quad (85)$$

(as  $\Phi_{\mathbf{y}}^{0:k} \geq 1$ ). Combining Assumption 2.5 with (84), (85), we get

$$|\Phi_\theta(x, y, F_{\theta, \mathbf{y}}^{0:k}(\mathcal{E}_\lambda)) - \Phi_\theta(x, y, F_{\theta, \mathbf{y}}^{m:k}(\Lambda))| \leq \varphi(x, y) \|F_{\theta, \mathbf{y}}^{0:k}(\mathcal{E}_\lambda) - F_{\theta, \mathbf{y}}^{m:k}(\Lambda)\| (\|F_{\theta, \mathbf{y}}^{0:k}(\mathcal{E}_\lambda)\| + \|F_{\theta, \mathbf{y}}^{m:k}(\Lambda)\|)^q \\ \leq \tilde{C}_4 \varphi(x, y) \tau^{2(k-m)} \|\Lambda\|^s (\Phi_{\mathbf{y}}^{0:k})^r. \quad (86)$$

Using (86), we deduce

$$\begin{aligned} & \max \left\{ |G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda)|, |G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda)| \right\} \\ & \leq \tilde{C}_4 \tau^{2(n-1)} n^{2r} \|\Lambda\|^s \varphi(X_n, Y_n) \sum_{k=1}^n \psi^r(Y_k) \leq C_5 \tau^n \|\Lambda\|^s \varphi(X_n, Y_n) \sum_{k=1}^n \psi^r(Y_k). \end{aligned}$$

Similarly, we conclude

$$\begin{aligned} & \max \left\{ |H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda)|, |H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda)| \right\} \\ & \leq \tilde{C}_4 \tau^{2(n-1)} n^{2r} \|\Lambda\|^s \varphi(X_{n+1}, Y_{n+1}) \sum_{k=1}^n \psi^r(Y_k) \leq C_5 \tau^n \|\Lambda\|^s \varphi(X_{n+1}, Y_{n+1}) \sum_{k=1}^n \psi^r(Y_k). \end{aligned}$$

□

**Lemma 6.2.** *Let Assumptions 2.1, 2.2 and 2.4 – 2.6 hold. Moreover, let  $\rho = \max\{\tau^{1/3}, \delta^{1/3}\}$  ( $\delta$  is specified in Assumption 2.4, while  $\tau$  is defined at the beginning of Section 5). Then, the following is true.*

(i) *There exists a real number  $C_6 \in [1, \infty)$  (depending only on  $p, q, \varepsilon, \delta, K_0, L_0$ ) such that*

$$\max \left\{ |A_\theta^n(x, \mathcal{E}_\lambda)|, |\tilde{A}_\theta^{m,n}(x, \mathcal{E}_\lambda)|, |\tilde{B}_\theta^n(x, \mathcal{E}_\lambda)| \right\} \leq C_6 \rho^{2n}, \quad |B_\theta^n(x, \mathcal{E}_\lambda) - B_\theta^n(x, \Lambda)| \leq C_6 \rho^{2n} \|\Lambda\|^s$$

for all  $\theta \in \Theta, x \in \mathcal{X}, \lambda \in \mathcal{P}(\mathcal{X}), \Lambda \in \mathcal{L}_0(\mathcal{X}), n > m \geq 0$ .

(ii) *There exists a real number  $C_7 \in [1, \infty)$  (depending only on  $p, q, \varepsilon, \delta, K_0, L_0$ ) such that*

$$|C_\theta^n(x, y, \mathcal{E}_\lambda)| \leq C_7 \rho^{2n} \psi^r(y), \quad |D_\theta^n(x, y, \mathcal{E}_\lambda) - D_\theta^n(x, y, \Lambda)| \leq C_7 \rho^{2n} \|\Lambda\|^s \psi^r(y)$$

for all  $\theta \in \Theta, x \in \mathcal{X}, y \in \mathcal{Y}, \lambda \in \mathcal{P}(\mathcal{X}), \Lambda \in \mathcal{L}_0(\mathcal{X}), n \geq 1$ .

*Proof.* Throughout the proof, the following notation is used.  $\tilde{C}_1, \tilde{C}_2$  are the real numbers defined by  $\tilde{C}_1 = \max_{n \geq 1} \rho^{n-1} n, \tilde{C}_2 = L_0^2$  ( $L_0$  is specified in Assumption 2.6).  $\theta, x, y, \lambda, \Lambda$  are any elements of  $\Theta, \mathcal{X}, \mathcal{Y}, \mathcal{P}(\mathcal{X}), \mathcal{L}_0(\mathcal{X})$  (respectively).  $n, m$  are any integers satisfying  $n > m \geq 0$ .

Owing to Assumption 2.5, we have

$$E(\varphi(X_k, Y_k) \psi^r(Y_k) | X_0 = x) = E\left(\int \varphi(X_k, y) \psi^r(y) Q(X_k, dy) \Big| X_0 = x\right) \leq L_0 \quad (87)$$

for  $k \geq 0$ . Due to the same assumption, we have

$$\max \left\{ \int \varphi(x, y') Q(x, dy'), \int \psi^r(y') Q(x, dy') \right\} \leq \int \varphi(x, y') \psi^r(y') Q(x, dy') \leq L_0. \quad (88)$$

Consequently, we get

$$E(\varphi(X_l, Y_l) \psi^r(Y_k) | X_0 = x) = E\left(\int \varphi(X_l, y) Q(X_l, dy) \int \psi^r(y) Q(X_k, dy) \Big| X_0 = x\right) \leq L_0^2 \quad (89)$$

for  $l > k \geq 0$ . Similarly, we get

$$E(\varphi(X_k, Y_k) \psi^r(Y_1) | X_1 = x, Y_1 = y) = \psi^r(y) E\left(\int \varphi(X_k, y') Q(X_k, dy') \Big| X_1 = x\right) \leq L_0 \psi^r(y), \quad (90)$$

$$E(\varphi(X_l, Y_l) \psi^r(Y_k) | X_1 = x, Y_1 = y) = E\left(\int \varphi(X_l, y') Q(X_l, dy') \int \psi^r(y') Q(X_k, dy') \Big| X_1 = x\right) \leq L_0^2 \quad (91)$$

for  $l > k > 1$ .

Let  $C_6$  be the real number defined by  $C_6 = \tilde{C}_1 \tilde{C}_2 C_5 K_0$  ( $K_0, C_5$  are specified in Assumption 2.4 and Lemma 6.1). Since  $\tau^n n \leq \rho^{3n}(n+1) \leq \tilde{C}_1 \rho^{2n}$ , Lemma 6.1 and (87), (89) imply

$$\begin{aligned} |A_\theta^n(x, \mathcal{E}_\lambda)| &\leq E(|G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\mathcal{E}_\lambda)| | X_0 = x) \\ &\leq C_5 \tau^n \sum_{k=1}^n E(|\varphi(X_n, Y_n) \psi^r(Y_k)| | X_0 = x) \\ &\leq \tilde{C}_2 C_5 \tau^n n \leq C_6 \rho^{2n}. \end{aligned}$$

As  $\tau^{n-m} \delta^m (n-m) \leq \rho^{3n}(n+1) \leq \tilde{C}_1 \rho^{2n}$ , Assumption 2.5 yields

$$|\tilde{A}_\theta^{m,n}(x, \mathcal{E}_\lambda)| \leq \int |A_\theta^{n-m}(x', \mathcal{E}_\lambda)| |P^m - \pi|(x, dx') \leq \tilde{C}_2 C_5 K_0 \tau^{n-m} \delta^m (n-m) \leq C_6 \rho^{2n}.$$

Moreover, owing to Lemma 6.1 and (87), (89), we have

$$\begin{aligned} |B_\theta^n(x, \mathcal{E}_\lambda) - B_\theta^n(x, \Lambda)| &\leq E(|G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda)| | X_0 = x) \\ &\leq C_5 \tau^n \|\Lambda\|^s \sum_{k=1}^n E(|\varphi(X_n, Y_n) \psi^r(Y_k)| | X_0 = x) \\ &\leq \tilde{C}_2 C_5 \tau^n n \|\Lambda\|^s \leq C_6 \rho^{2n} \|\Lambda\|^s. \end{aligned}$$

Similarly, due to Assumptions 2.4, 2.5 and (88), we have

$$\begin{aligned} |\tilde{B}_\theta^n(x, \mathcal{E}_\lambda)| &\leq \int \int |\Phi_\theta(x', y', \mathcal{E}_\lambda)| Q(x', dy') |P^n - \pi|(x, dx') \\ &\leq \int \int \varphi(x', y') Q(x', dy') |P^n - \pi|(x, dx') \\ &\leq \tilde{C}_2 K_0 \delta^n \leq C_6 \rho^{2n}. \end{aligned}$$

Let  $C_7$  be the real number defined by  $C_7 = \tilde{C}_1 \tilde{C}_2 C_5$  ( $C_5$  is specified in Lemma 6.1). Relying on Lemma 6.1 and (90), (91), we deduce

$$\begin{aligned} |C_\theta^m(x, y, \mathcal{E}_\lambda)| &\leq E(|H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\mathcal{E}_\lambda)| | X_1 = x, Y_1 = y) \\ &\leq C_5 \tau^n \sum_{k=1}^n E(|\varphi(X_{n+1}, Y_{n+1}) \psi^r(Y_k)| | X_1 = x, Y_1 = y) \\ &\leq \tilde{C}_2 C_5 \tau^n n \psi^r(y) \leq C_7 \rho^{2n} \psi^r(y). \end{aligned}$$

Using the same arguments, we conclude

$$\begin{aligned} |D_\theta^n(x, y, \mathcal{E}_\lambda) - D_\theta^n(x, y, \Lambda)| &\leq E(|H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda)| | X_1 = x, Y_1 = y) \\ &\leq C_5 \tau^n \|\Lambda\|^s \sum_{k=1}^n E(|\varphi(X_{n+1}, Y_{n+1}) \psi^r(Y_k)| | X_1 = x, Y_1 = y) \\ &\leq \tilde{C}_2 C_5 \tau^n n \|\Lambda\|^s \psi^r(y) \leq C_7 \rho^{2n} \|\Lambda\|^s \psi^r(y). \end{aligned}$$

□

**Proof of Theorem 2.3.** Throughout the proof, the following notation is used.  $\tilde{C}_1$  is the real number defined by  $\tilde{C}_1 = \max_{n \geq 1} \rho^{n-1} n$ , while  $\tilde{C}_2, \tilde{C}_3$  are the real numbers defined by  $\tilde{C}_2 = 4\tilde{C}_1 C_6$ ,  $\tilde{C}_3 = \tilde{C}_2(1-\rho)^{-1}$  ( $\rho, C_6$  are specified in Lemma 6.2).  $L$  is the real number defined by  $L = 4\tilde{C}_3 C_7 L_0 \rho^{-1}$  ( $L_0, C_7$  are specified in Assumption 2.6 and Lemma 6.2).  $\theta$  is any element of  $\Theta$ .  $x, x'$  are any elements of  $\mathcal{X}$ , while  $y, y'$  are any elements of  $\mathcal{Y}$ .  $\lambda, \lambda'$  are any elements of  $\mathcal{P}(\mathcal{X})$ , while  $\Lambda, \Lambda'$  are any elements of  $\mathcal{L}_0(\mathcal{X})$ .  $n$  is any (strictly) positive integer.

It is easy to notice that  $G_{\theta, \mathbf{X}, \mathbf{Y}}^{l:n}(\mathcal{E}_\lambda)$  does not depend on  $X_0, Y_0, \dots, X_k, Y_k$  for  $n \geq l \geq k \geq 0$ . It is also easy to show

$$E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{l:n}(\mathcal{E}_\lambda) \middle| X_k = x \right) = E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{l-k:n-k}(\mathcal{E}_\lambda) \middle| X_0 = x \right)$$

for the same  $k, l$ . Then, we conclude

$$\begin{aligned} (\Pi^n \Phi)_\theta(x, y, \Lambda) &= E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) \middle| X_0 = x \right) = \sum_{k=0}^{n-1} E \left( E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{k:n}(\mathcal{E}_\lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{k+1:n}(\mathcal{E}_\lambda) \middle| X_k \right) \middle| X_0 = x \right) \\ &\quad + E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) \middle| X_0 = x \right) \\ &\quad + E \left( E \left( \Phi_\theta(X_n, Y_n, \mathcal{E}_\lambda) \middle| X_n \right) \middle| X_0 = x \right) \\ &= \sum_{k=0}^{n-1} \left( \tilde{A}_\theta^{k:n}(x, \mathcal{E}_\lambda) + \bar{A}_\theta^{k:n}(\mathcal{E}_\lambda) \right) + B_\theta^n(x, \Lambda) - B_\theta^n(x, \mathcal{E}_\lambda) \\ &\quad + \tilde{B}_\theta^n(x, \mathcal{E}_\lambda) + \bar{B}_\theta^n(\mathcal{E}_\lambda), \end{aligned} \quad (92)$$

where

$$\bar{A}_\theta^{k,n}(\mathcal{E}_\lambda) = \int A_\theta^{n-k}(x', \mathcal{E}_\lambda) \pi(dx'), \quad \bar{B}_\theta^n(\mathcal{E}_\lambda) = \int \int \Phi_\theta^n(x', y', \mathcal{E}_\lambda) Q(x', dy') \pi(dx').$$

We also deduce

$$\begin{aligned} (\Pi^n \Phi)_\theta(x, y, \Lambda) &= E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) \middle| X_0 = x \right) = E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) - G_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda) \middle| X_0 = x \right) \\ &\quad + E \left( E \left( G_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\Lambda) \middle| X_1 \right) \middle| X_0 = x \right) \\ &= A_\theta^n(x, \Lambda) + E \left( (\Pi^{n-1} \Phi)_\theta(X_1, Y_1, \Lambda) \middle| X_0 = x \right). \end{aligned} \quad (93)$$

Since  $\rho^{2n}(n+1) \leq \tilde{C}_1 \rho^n$ , Lemma 6.2 and (92) imply

$$\begin{aligned} |(\Pi^n \Phi)_\theta(x, y, \Lambda) - (\Pi^n \Phi)_\theta(x', y', \mathcal{E}_\lambda)| &\leq \left| \tilde{B}_\theta^n(x, \mathcal{E}_\lambda) \right| + \left| \tilde{B}_\theta^n(x', \mathcal{E}_\lambda) \right| + |B_\theta^n(x, \Lambda) - B_\theta^n(x, \mathcal{E}_\lambda)| \\ &\quad + \sum_{k=0}^{n-1} \left| \tilde{A}_\theta^{k:n}(x, \mathcal{E}_\lambda) \right| + \sum_{k=0}^{n-1} \left| \tilde{A}_\theta^{k:n}(x', \mathcal{E}_\lambda) \right| \\ &\leq 2C_6 \rho^{2n}(n+1) + C_6 \rho^{2n} \|\Lambda\|^s \leq \tilde{C}_2 \rho^n \|\Lambda\|^s. \end{aligned} \quad (94)$$

Then, Lemma 6.2 and (93) yield

$$\begin{aligned} |(\Pi^{n+1} \Phi)_\theta(x, y, \mathcal{E}_\lambda) - (\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda)| &\leq E \left( |(\Pi^n \Phi)_\theta(X_1, Y_1, \mathcal{E}_\lambda) - (\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda)| \middle| X_0 = x \right) \\ &\quad + |A_\theta^{n+1}(x, \mathcal{E}_\lambda)| \\ &\leq 2C_6 \rho^{2n}(n+2) + C_6 \rho^{2(n+1)} \leq \tilde{C}_2 \rho^n. \end{aligned} \quad (95)$$

Let  $\phi_\theta(x, y, \mathcal{E}_\lambda)$  be the function defined by

$$\phi_\theta(x, y, \mathcal{E}_\lambda) = \Phi_\theta(x, y, \mathcal{E}_\lambda) + \sum_{n=0}^{\infty} \left( (\Pi^{n+1} \Phi)_\theta(x, y, \mathcal{E}_\lambda) - (\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda) \right).$$

Owing to (95),  $\phi_\theta(x, y, \mathcal{E}_\lambda)$  is well-defined. Due to the same inequality, we have

$$|(\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda) - \phi_\theta(x, y, \mathcal{E}_\lambda)| \leq \sum_{k=n}^{\infty} |(\Pi^{k+1} \Phi)_\theta(x, y, \mathcal{E}_\lambda) - (\Pi^k \Phi)_\theta(x, y, \mathcal{E}_\lambda)| \leq \tilde{C}_2 \sum_{k=n}^{\infty} \rho^k = \tilde{C}_3 \rho^n. \quad (96)$$

Consequently, (94) yields

$$\begin{aligned} |\phi_\theta(x, y, \mathcal{E}_\lambda) - \phi_\theta(x', y', \mathcal{E}_{\lambda'})| &\leq |(\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda) - (\Pi^n \Phi)_\theta(x', y', \mathcal{E}_{\lambda'})| \\ &\quad + |(\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda) - \phi_\theta(x, y, \mathcal{E}_\lambda)| \\ &\quad + |(\Pi^n \Phi)_\theta(x', y', \mathcal{E}_{\lambda'}) - \phi_\theta(x', y', \mathcal{E}_{\lambda'})| \\ &\leq (\tilde{C}_2 + 2\tilde{C}_3) \rho^n. \end{aligned}$$

Letting  $n \rightarrow \infty$ , we conclude  $\phi_\theta(x, y, \mathcal{E}_\lambda) = \phi_\theta(x', y', \mathcal{E}_{\lambda'})$ . Hence, there exists a function  $\phi_\theta$  which maps  $\theta$  to  $\mathbb{R}$  and satisfies  $\phi_\theta = \phi_\theta(x, y, \mathcal{E}_\lambda)$  for each  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ . Then, (94), (96) imply

$$\begin{aligned} |(\Pi^n \Phi)_\theta(x, y, \Lambda) - \phi_\theta| &\leq |(\Pi^n \Phi)_\theta(x, y, \Lambda) - (\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda)| + |(\Pi^n \Phi)_\theta(x, y, \mathcal{E}_\lambda) - \phi_\theta| \\ &\leq \tilde{C}_2 \rho^n \|\Lambda\|^s + \tilde{C}_3 \rho^n \leq 2\tilde{C}_3 \rho^n \|\Lambda\|^s \leq L \rho^n \|\Lambda\|^s \end{aligned} \quad (97)$$

(as  $\|\Lambda\| \geq 1$ ).

Owing to Assumption 2.5, we have

$$\left| \tilde{\Phi}_\theta(x, y, \Lambda) \right| \leq \int \int |\Phi_\theta(x', y', \Lambda)| Q(x', dy') P(x, dx') \leq \int \int \varphi(x', y') \|\Lambda\|^q Q(x', dy') P(x, dx') \leq L_0 \|\Lambda\|^q \quad (98)$$

(see also (88)). Due to the same assumption, we have

$$\begin{aligned} \left| \tilde{\Phi}_\theta(x, y, \Lambda) - \tilde{\Phi}_\theta(x, y, \Lambda') \right| &\leq \int \int |\Phi_\theta(x', y', \Lambda) - \Phi_\theta(x', y', \Lambda')| Q(x', dy') P(x, dx') \\ &\leq \int \int \varphi(x', y') \|\Lambda - \Lambda'\| (\|\Lambda\| + \|\Lambda'\|)^q Q(x', dy') P(x, dx') \\ &\leq L_0 \|\Lambda - \Lambda'\| (\|\Lambda\| + \|\Lambda'\|)^q. \end{aligned} \quad (99)$$

Using (98), (99), we conclude that Assumption 2.5 holds when  $\Phi_\theta(x, y, \Lambda)$  is replaced by  $\tilde{\Phi}_\theta(x, y, \Lambda)/L_0$ . Consequently, Assumption 2.5 and (97) imply that there exists a function  $\tilde{\phi}_\theta$  mapping  $\theta$  to  $\mathbb{R}$  such that (97) is still true when  $\Phi_\theta(x, y, \Lambda)$ ,  $\phi_\theta$  are replaced with  $\tilde{\Phi}_\theta(x, y, \Lambda)/L_0$ ,  $\tilde{\phi}_\theta/L_0$  (respectively). Hence, we get

$$\left| (\Pi^n \tilde{\Phi})_\theta(x, y, \Lambda) - \tilde{\phi}_\theta \right| \leq 2\tilde{C}_3 L_0 \rho^n \|\Lambda\|^s. \quad (100)$$

Moreover, it is easy notice that  $H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\mathcal{E}_\lambda)$  does not depend on  $X_1, Y_1, X_2, Y_2$ . Then, we conclude

$$\begin{aligned} (\tilde{\Pi}^n \Phi)_\theta(x, y, \Lambda) &= E \left( H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) \mid X_1 = x, Y_1 = y \right) = E \left( H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\Lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) \mid X_1 = x, Y_1 = y \right) \\ &\quad + E \left( H_{\theta, \mathbf{X}, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda) - H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\mathcal{E}_\lambda) \mid X_1 = x, Y_1 = y \right) \\ &\quad + E \left( E \left( H_{\theta, \mathbf{X}, \mathbf{Y}}^{1:n}(\mathcal{E}_\lambda) \mid X_{2:n}, Y_{2:n} \right) \mid X_1 = x, Y_1 = y \right) \\ &= C_\theta^n(x, y, \mathcal{E}_\lambda) + D_\theta^n(x, y, \Lambda) - D_\theta^n(x, y, \mathcal{E}_\lambda) \\ &\quad + E \left( \tilde{\Phi}_\theta(X_n, Y_n, F_{\theta, \mathbf{Y}}^{1:n}(\mathcal{E}_\lambda)) \mid X_1 = x, Y_1 = y \right) \\ &= C_\theta^n(x, y, \mathcal{E}_\lambda) + D_\theta^n(x, y, \Lambda) - D_\theta^n(x, y, \mathcal{E}_\lambda) \\ &\quad + (\Pi^{n-1} \tilde{\Phi})_\theta(x, y, \mathcal{E}_\lambda). \end{aligned}$$

Combining this with Lemma 6.2 and (100), we get

$$\begin{aligned} \left| (\tilde{\Pi}^n \Phi)_\theta(x, y, \Lambda) - \tilde{\phi}_\theta \right| &\leq \left| (\Pi^{n-1} \tilde{\Phi})_\theta(x, y, \mathcal{E}_\lambda) - \tilde{\phi}_\theta \right| + |C_\theta^n(x, y, \mathcal{E}_\lambda)| + |D_\theta^n(x, y, \Lambda) - D_\theta^n(x, y, \mathcal{E}_\lambda)| \\ &\leq 2\tilde{C}_3 L_0 \rho^{n-1} + C_7 \rho^{2n} \psi^r(y) + C_7 \rho^{2n} \psi^r(y) \|\Lambda\|^s \\ &\leq 4\tilde{C}_3 C_7 L_0 \rho^{n-1} \psi^r(y) \|\Lambda\|^s \leq L \rho^n \psi^r(y) \|\Lambda\|^s. \end{aligned}$$

□

## 7. Proof of Theorems 2.1 and 3.1

In this section, we rely on the following notation. For  $1 \leq i \leq d$ ,  $e_i$  denotes the  $i$ -th standard unit vector in  $\mathbb{N}_0^d$ .  $\mathbf{e}_\alpha$  is the vector defined by

$$i(\alpha) = \min\{i : e_i \leq \alpha, 1 \leq i \leq d\}, \quad \mathbf{e}_\alpha = e_{i(\alpha)}$$

for  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ .  $\Psi_\theta(y, \lambda)$ ,  $\Psi_\theta^0(y, \Lambda)$  and  $\Psi_\theta^\alpha(y, \Lambda)$  are the functions defined by

$$\Psi_\theta(y, \lambda) = \log \langle R_{\theta, y}^0(\lambda) \rangle, \quad \Psi_\theta^0(y, \Lambda) = \Psi_\theta(y, \lambda_0), \quad \Psi_\theta^\alpha(y, \Lambda) = \langle S_{\theta, y}^\alpha(\Lambda) \rangle \quad (101)$$

for  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $\Lambda = \{\lambda_\beta : \beta \in \mathbb{N}_0^d, |\beta| \leq p\} \in \mathcal{L}_0(\mathcal{X})$  and  $|\alpha| = 1$ .  $\Psi_\theta^\alpha(y, \Lambda)$  is the function recursively defined by

$$\Psi_\theta^\alpha(y, \Lambda) = \langle S_{\theta, y}^\alpha(\Lambda) \rangle - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} \Psi_\theta^\beta(y, \Lambda) \langle S_{\theta, y}^{\alpha - \beta}(\Lambda) \rangle \quad (102)$$

for  $1 < |\alpha| \leq p$ , where the recursion is in  $|\alpha|$ .<sup>4</sup>

**Proposition 7.1.** *Let Assumptions 2.1 – 2.3 hold. Then,  $p_{\theta, \mathbf{y}}^{0:n}(x|\lambda)$ ,  $P_{\theta, \mathbf{y}}^{0:n}(B|\lambda)$  and  $\Psi_\theta^0(y_{n+1}, P_{\theta, \mathbf{y}}^{0:n}(\lambda))$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $B \in \mathcal{B}(\mathcal{X})$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $n \geq 1$  and any sequence  $\mathbf{y} = \{y_n\}_{n \geq 1}$  in  $\mathcal{Y}$  ( $y_{n+1}$  is the  $(n+1)$ -th element of  $\mathbf{y}$ ). Moreover, we have*

$$\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{0:n}(x|\lambda) = f_{\theta, \mathbf{y}}^{\alpha, 0:n}(x|\mathcal{E}_\lambda), \quad \partial_\theta^\alpha P_{\theta, \mathbf{y}}^{0:n}(B|\lambda) = F_{\theta, \mathbf{y}}^{\alpha, 0:n}(B|\mathcal{E}_\lambda), \quad \partial_\theta^\alpha \Psi_\theta^0(y_{n+1}, P_{\theta, \mathbf{y}}^{0:n}(\lambda)) = \Psi_\theta^\alpha(y_{n+1}, F_{\theta, \mathbf{y}}^{\alpha, 0:n}(\mathcal{E}_\lambda)) \quad (103)$$

for the same  $\theta$ ,  $x$ ,  $B$ ,  $\lambda$ ,  $n$ ,  $\mathbf{y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$  ( $\mathcal{E}_\lambda$ ,  $f_{\theta, \mathbf{y}}^{\alpha, 0:n}(x|\mathcal{E}_\lambda)$ ,  $p_{\theta, \mathbf{y}}^{0:n}(x|\lambda)$ ,  $F_{\theta, \mathbf{y}}^{\alpha, 0:n}(B|\mathcal{E}_\lambda)$ ,  $P_{\theta, \mathbf{y}}^{0:n}(B|\lambda)$  are defined in (3), (10) – (12)).

*Proof.* Throughout the proof, the following notation is used.  $\theta$ ,  $\lambda$ ,  $B$  are any elements of  $\Theta$ ,  $\mathcal{P}(\mathcal{X})$ ,  $\mathcal{B}(\mathcal{X})$  (respectively), while  $x, x'$  are any elements of  $\mathcal{X}$ .  $\mathbf{y} = \{y_n\}_{n \geq 1}$  is any sequence in  $\mathcal{Y}$ , while  $\alpha$  is any element of  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .  $n$  is any (strictly) positive integer.  $\delta_x(dx')$  is the Dirac measure centered at  $x$ .  $\xi_n(dx_{0:n}|x, \lambda)$  and  $\zeta(dx_{0:n}|\lambda)$  are the measures on  $\mathcal{X}^{n+1}$  defined by

$$\xi_n(A|x, \lambda) = \int \int \cdots \int I_A(x_{0:n}) \delta_x(dx_n) \mu(dx_{n-1}) \cdots \mu(dx_1) \lambda(dx_0), \quad (104)$$

$$\zeta_n(A|\lambda) = \int \cdots \int \int I_A(x_{0:n}) \mu(dx_n) \cdots \mu(dx_1) \lambda(dx_0) \quad (105)$$

for  $A \in \mathcal{B}(\mathcal{X}^{n+1})$ .<sup>5</sup>  $u_{\theta, \mathbf{y}}^n(x_{0:n})$  is the function defined by

$$u_{\theta, \mathbf{y}}^n(x_{0:n}) = \prod_{k=1}^n r_\theta(y_k, x_k | x_{k-1})$$

for  $x_0, \dots, x_n \in \mathcal{X}$ .  $v_{\theta, \mathbf{y}}^n(x|\lambda)$  and  $w_{\theta, \mathbf{y}}^n(\lambda)$  are the functions defined by

$$v_{\theta, \mathbf{y}}^n(x|\lambda) = \int u_{\theta, \mathbf{y}}^n(x_{0:n}) \xi_n(dx_{0:n}|x, \lambda), \quad w_{\theta, \mathbf{y}}^n(\lambda) = \int u_{\theta, \mathbf{y}}^n(x_{0:n}) \zeta_n(dx_{0:n}|\lambda). \quad (106)$$

Using (3), it is straightforward to verify

$$p_{\theta}^{0:n}(x|\lambda) = \frac{v_{\theta, \mathbf{y}}^n(x|\lambda)}{w_{\theta, \mathbf{y}}^n(\lambda)}, \quad P_{\theta}^{0:n}(B|\lambda) = \int_B \frac{v_{\theta, \mathbf{y}}^n(x'|\lambda)}{w_{\theta, \mathbf{y}}^n(\lambda)} \mu(dx'), \quad w_{\theta, \mathbf{y}}^n(\lambda) = \int v_{\theta, \mathbf{y}}^n(x'|\lambda) \mu(dx'). \quad (107)$$

It is also easy to show

$$w_{\theta, \mathbf{y}}^1(\lambda) = \int \left( \int r_\theta(y_1, x' | x) \mu(dx') \right) \lambda(dx), \quad w_{\theta, \mathbf{y}}^{n+1}(\lambda) = \int \left( \int r_\theta(y_{n+1}, x' | x) \mu(dx') \right) v_{\theta, \mathbf{y}}^n(x|\lambda) \mu(dx).$$

<sup>4</sup>The last two functions in (101) are initial conditions in (102). At iteration  $k$  of (102) ( $1 < k \leq p$ ), function  $\Psi_\theta^\alpha(y, \Lambda)$  is computed for multi-indices  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = k$  using the results obtained at the previous iterations.

<sup>5</sup>When  $n = 1$ , (104), (105) should be interpreted as

$$\xi_1(A|x, \lambda) = \int \int I_A(x_{0:1}) \delta_x(dx_1) \lambda(dx_0), \quad \zeta_1(A|\lambda) = \int \int I_A(x_{0:1}) \mu(dx_1) \lambda(dx_0).$$



Consequently, Assumption 2.2 implies

$$\begin{aligned} w_{\theta, \mathbf{y}}^1(\lambda) &\geq \varepsilon \int \mu_{\theta}(\mathcal{X}|y_1) \lambda(dx) = \varepsilon \mu_{\theta}(\mathcal{X}|y_1), \\ w_{\theta, \mathbf{y}}^{n+1}(\lambda) &\geq \varepsilon \int \mu_{\theta}(\mathcal{X}|y_{n+1}) v_{\theta, \mathbf{y}}^n(x|\lambda) \mu(dx) = \varepsilon \mu_{\theta}(\mathcal{X}|y_{n+1}) w_{\theta, \mathbf{y}}^n(\lambda). \end{aligned} \quad (108)$$

Iterating (108), we get

$$w_{\theta, \mathbf{y}}^n(\lambda) \geq \varepsilon^n \prod_{k=1}^n \mu_{\theta}(\mathcal{X}|y_k) > 0. \quad (109)$$

Owing to Leibniz rule and Assumptions 2.2, 2.3, we have

$$\begin{aligned} |\partial_{\theta}^{\alpha} u_{\theta, \mathbf{y}}^n(x_{0:n})| &\leq \sum_{\substack{\beta_1, \dots, \beta_n \in \mathbb{N}_0^d \\ \beta_1 + \dots + \beta_n = \alpha}} \binom{\alpha}{\beta_1, \dots, \beta_n} \prod_{k=1}^n \left| \partial_{\theta}^{\beta_k} r_{\theta}(y_k, x_k | x_{k-1}) \right| \\ &\leq \left( \prod_{k=1}^n \phi(y_k, x_k) \right) \left( \sum_{\substack{\beta_1, \dots, \beta_n \in \mathbb{N}_0^d \\ \beta_1 + \dots + \beta_n = \alpha}} \binom{\alpha}{\beta_1, \dots, \beta_n} \prod_{k=1}^n (\psi(y_k))^{\beta_k} \right) \\ &\leq 2^{|\alpha|} \left( \prod_{k=1}^n \psi(y_k) \right)^{|\alpha|} \left( \prod_{k=1}^n \phi(y_k, x_k) \right) \end{aligned} \quad (110)$$

for  $x_0, \dots, x_n \in \mathcal{X}$ . Due to the same assumptions, we have

$$\int \left( \prod_{k=1}^n \phi(y_k, x_k) \right) \xi_n(dx_{0:n} | x, \lambda) = \phi(y_n, x) \left( \prod_{k=1}^{n-1} \int \phi(y_k, x_k) \mu(dx_k) \right) < \infty, \quad (111)$$

$$\int \left( \prod_{k=1}^n \phi(y_k, x_k) \right) \zeta_n(dx_{0:n} | \lambda) = \left( \prod_{k=1}^n \int \phi(y_k, x_k) \mu(dx_k) \right) < \infty. \quad (112)$$

Here and throughout the proof, we rely on the convention that  $\prod_{k=i}^j$  is equal to one whenever  $j < i$ . Using Lemma A3.1 (see Appendix 3) and (110) – (112), we conclude that  $v_{\theta, \mathbf{y}}^n(x|\lambda)$ ,  $w_{\theta, \mathbf{y}}^n(\lambda)$  are well-defined and  $p$ -times differentiable in  $\theta$ . Relying on the same arguments, we deduce

$$\partial_{\theta}^{\alpha} v_{\theta, \mathbf{y}}^n(x|\lambda) = \int \partial_{\theta}^{\alpha} u_{\theta, \mathbf{y}}^n(x_{0:n}) \xi_n(dx_{0:n} | x, \lambda), \quad \partial_{\theta}^{\alpha} w_{\theta, \mathbf{y}}^n(\lambda) = \int \partial_{\theta}^{\alpha} u_{\theta, \mathbf{y}}^n(x_{0:n}) \zeta_n(dx_{0:n} | \lambda). \quad (113)$$

Then, (107), (109) imply that  $p_{\theta, \mathbf{y}}^{0:n}(x|\lambda)$  is  $p$ -times differentiable in  $\theta$ . Moreover, (110), (113) yield

$$|\partial_{\theta}^{\alpha} v_{\theta, \mathbf{y}}^n(x|\lambda)| \leq \int |\partial_{\theta}^{\alpha} u_{\theta, \mathbf{y}}^n(x_{0:n})| \xi_n(dx_{0:n} | x, \lambda) \leq 2^{|\alpha|} \phi(y_n, x) \left( \prod_{k=1}^n \psi(y_k) \right)^{|\alpha|} \left( \prod_{k=1}^{n-1} \int \phi(y_k, x_k) \mu(dx_k) \right). \quad (114)$$

Let  $\tilde{P}_{\theta, \mathbf{y}}^{\alpha, n}(dx|\lambda)$  be the signed measure defined by

$$\tilde{P}_{\theta, \mathbf{y}}^{\alpha, n}(B|\lambda) = \int_B \partial_{\theta}^{\alpha} p_{\theta, \mathbf{y}}^{0:n}(x|\lambda), \mu(dx) \quad (115)$$

while  $\tilde{P}_{\theta, \mathbf{y}}^{\alpha, n}(\lambda)$  is a ‘short-hand’ notation for  $\tilde{P}_{\theta, \mathbf{y}}^{\alpha, n}(dx|\lambda)$ . Moreover, let  $\tilde{P}_{\theta, \mathbf{y}}^0(\lambda)$  and  $\tilde{P}_{\theta, \mathbf{y}}^n(\lambda)$  be the vector measures defined by

$$\tilde{P}_{\theta, \mathbf{y}}^0(\lambda) = \mathcal{E}_{\lambda}, \quad \tilde{P}_{\theta, \mathbf{y}}^n(\lambda) = \left\{ \tilde{P}_{\theta, \mathbf{y}}^{\alpha, n}(\lambda) : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p \right\}, \quad (116)$$

where  $\tilde{P}_{\theta,\mathbf{y}}^{\alpha,n}(\lambda)$  is the component  $\alpha$  of  $\tilde{P}_{\theta,\mathbf{y}}^n(\lambda)$ . Owing to Lemma A3.1 and (107), (109), (114),  $P_{\theta,\mathbf{y}}^{0:n}(B|\lambda)$  is  $p$ -times differentiable in  $\theta$ . Due to the same arguments,  $\tilde{P}_{\theta,\mathbf{y}}^{\alpha,n}(B|\lambda)$  is well-defined and satisfies

$$\tilde{P}_{\theta,\mathbf{y}}^{\alpha,n}(B|\lambda) = \partial_{\theta}^{\alpha} P_{\theta,\mathbf{y}}^{0:n}(B|\lambda) = \partial_{\theta}^{\alpha} \tilde{P}_{\theta,\mathbf{y}}^{0,n}(B|\lambda) \quad (117)$$

(as  $P_{\theta,\mathbf{y}}^{0:n}(\lambda) = \tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)$ ).

Using (5), (8), (107), (115), it is straightforward to verify

$$r_{\theta,y_{n+1}}^{\mathbf{0}}(x|\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) = \frac{\int r_{\theta}(y_{n+1}, x|x') v_{\theta,\mathbf{y}}^n(x'|\lambda) \mu(dx')}{w_{\theta,\mathbf{y}}^n(\lambda)}, \quad (118)$$

$$\left\langle R_{\theta,y_{n+1}}^{\mathbf{0}}(\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) \right\rangle = \frac{\int \int r_{\theta}(y_{n+1}, x''|x') v_{\theta,\mathbf{y}}^n(x'|\lambda) \mu(dx') \mu(dx'')}{w_{\theta,\mathbf{y}}^n(\lambda)}. \quad (119)$$

Moreover, Leibniz rule, Assumptions 2.2, 2.3 and (114) imply

$$\begin{aligned} |\partial_{\theta}^{\alpha} (r_{\theta}(y_{n+1}, x|x') v_{\theta,\mathbf{y}}^n(x'|\lambda))| &\leq \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} |\partial_{\theta}^{\alpha-\beta} r_{\theta}(y_{n+1}, x|x')| |\partial_{\theta}^{\beta} v_{\theta,\mathbf{y}}^n(x'|\lambda)| \\ &\leq \phi(y_{n+1}, x) \phi(y_n, x') \left( \prod_{k=1}^{n-1} \int \phi(y_k, x_k) \mu(dx_k) \right) \left( \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} 2^{|\beta|} (\psi(y_{n+1}))^{|\alpha-\beta|} \left( \prod_{k=1}^n \psi(y_k) \right)^{|\beta|} \right) \\ &\leq 4^{|\alpha|} \phi(y_{n+1}, x) \phi(y_n, x') \left( \prod_{k=1}^{n+1} \psi(y_k) \right)^{|\alpha|} \left( \prod_{k=1}^{n-1} \int \phi(y_k, x_k) \mu(dx_k) \right) < \infty. \end{aligned} \quad (120)$$

The same assumptions also yield

$$\int \phi(y_n, x') \mu(dx') < \infty, \quad \int \int \phi(y_{n+1}, x) \phi(y_n, x') \mu(dx) \mu(dx') < \infty. \quad (121)$$

Using Lemma A3.1 and (109), (118) – (121), we conclude that  $r_{\theta,y_{n+1}}^{\mathbf{0}}(x|\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda))$ ,  $\langle R_{\theta,y_{n+1}}^{\mathbf{0}}(\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) \rangle$  are well-defined and  $p$ -times differentiable in  $\theta$ .<sup>6</sup> Relying on the same arguments and (107), we deduce

$$\partial_{\theta}^{\alpha} r_{\theta,y_{n+1}}^{\mathbf{0}}(x|\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) = \int \partial_{\theta}^{\alpha} (r_{\theta}(y_{n+1}, x|x') p_{\theta,\mathbf{y}}^{0:n}(x'|\lambda)) \mu(dx'), \quad (122)$$

$$\partial_{\theta}^{\alpha} \left\langle R_{\theta,y_{n+1}}^{\mathbf{0}}(\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) \right\rangle = \int \int \partial_{\theta}^{\alpha} (r_{\theta}(y_{n+1}, x''|x') p_{\theta,\mathbf{y}}^{0:n}(x'|\lambda)) \mu(dx'') \mu(dx'). \quad (123)$$

Consequently, Leibniz rule and (5), (8), (115) imply

$$\begin{aligned} \partial_{\theta}^{\alpha} r_{\theta,y_{n+1}}^{\mathbf{0}}(x|\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \int \partial_{\theta}^{\alpha-\beta} r_{\theta}(y_{n+1}, x|x') \partial_{\theta}^{\beta} p_{\theta,\mathbf{y}}^{0:n}(x'|\lambda) \mu(dx') \\ &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} r_{\theta,y_{n+1}}^{\alpha-\beta}(x|\tilde{P}_{\theta,\mathbf{y}}^{\beta,n}(\lambda)) \\ &= s_{\theta,y_{n+1}}^{\alpha}(x|\tilde{P}_{\theta,\mathbf{y}}^n(\lambda)) \left\langle R_{\theta,y_{n+1}}^{\mathbf{0}}(\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) \right\rangle. \end{aligned} \quad (124)$$

<sup>6</sup>To conclude that  $r_{\theta,y_{n+1}}^{\mathbf{0}}(x|\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda))$  is well-defined and satisfy (122), set  $z = x'$ ,  $\nu(dz) = \mu(dx')$ ,  $F_{\theta}(z) = r_{\theta}(y_{n+1}, x|x') v_{\theta,\mathbf{y}}^n(x'|\lambda)$ ,  $g_{\theta} = w_{\theta,\mathbf{y}}^n(\lambda)$  in Lemma A3.1 ( $x$  is treated as a fixed value). To conclude that  $\langle R_{\theta,y_{n+1}}^{\mathbf{0}}(\tilde{P}_{\theta,\mathbf{y}}^{0,n}(\lambda)) \rangle$  is well-defined and satisfy (123), set  $z = (x, x')$ ,  $\nu(dz) = \mu(dx) \mu(dx')$ ,  $F_{\theta}(z) = r_{\theta}(y_{n+1}, x|x') v_{\theta,\mathbf{y}}^n(x'|\lambda)$ ,  $g_{\theta} = w_{\theta,\mathbf{y}}^n(\lambda)$  in Lemma A3.1.

Leibniz rule and (5), (8), (115) also yield

$$\begin{aligned}
\partial_\theta^\alpha \langle R_{\theta, y_{n+1}}^0 (\tilde{P}_{\theta, \mathbf{y}}^{0, n}(\lambda)) \rangle &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \int \int \partial_\theta^{\alpha-\beta} r_\theta(y_{n+1}, x''|x') \partial_\theta^\beta p_{\theta, \mathbf{y}}^{0, n}(x'|\lambda) \mu(dx') \mu(dx'') \\
&= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \langle R_{\theta, y_{n+1}}^{\alpha-\beta} (\tilde{P}_{\theta, \mathbf{y}}^{\beta, n}(\lambda)) \rangle \\
&= \langle S_{\theta, y_{n+1}}^\alpha (\tilde{P}_{\theta, \mathbf{y}}^n(\lambda)) \rangle \langle R_{\theta, y_{n+1}}^0 (\tilde{P}_{\theta, \mathbf{y}}^{0, n}(\lambda)) \rangle.
\end{aligned} \tag{125}$$

Moreover, using (5), (8), (106), (115), we get

$$\begin{aligned}
\partial_\theta^\alpha r_{\theta, y_1}^0(x|\tilde{P}_{\theta, \mathbf{y}}^{0, 0}(\lambda)) &= \partial_\theta^\alpha v_{\theta, \mathbf{y}}^1(x|\lambda) = \int \partial_\theta^\alpha r_\theta(x, y_1|x') \lambda(dx') \\
&= s_{\theta, y_1}^\alpha(x|\tilde{P}_{\theta, \mathbf{y}}^0(\lambda)) \langle R_{\theta, y_1}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, 0}(\lambda)) \rangle,
\end{aligned} \tag{126}$$

$$\begin{aligned}
\partial_\theta^\alpha \langle R_{\theta, y_1}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, 0}(\lambda)) \rangle &= \partial_\theta^\alpha w_{\theta, \mathbf{y}}^1(\lambda) = \int \int \partial_\theta^\alpha r_\theta(x'', y_1|x') \mu(dx'') \lambda(dx') \\
&= \langle S_{\theta, y_1}^\alpha(\tilde{P}_{\theta, \mathbf{y}}^0(\lambda)) \rangle \langle R_{\theta, y_1}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, 0}(\lambda)) \rangle
\end{aligned} \tag{127}$$

(as  $\tilde{P}_{\theta, \mathbf{y}}^{0, 0}(B|\lambda) = \lambda(B)$ ).

Relying on (3), (5), (8), (115), it is straightforward to verify

$$p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle = r_{\theta, y_n}^0(x|\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)).$$

Then, Leibniz rule and (125), (127) imply

$$\begin{aligned}
\partial_\theta^\alpha r_{\theta, y_n}^0(x|\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \partial_\theta^\beta p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) \partial_\theta^{\alpha-\beta} \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle \\
&= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \partial_\theta^\beta p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) \langle S_{\theta, y_n}^{\alpha-\beta}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle.
\end{aligned}$$

Since  $0 < \langle R_{\theta, y_{n+1}}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n}(\lambda)) \rangle < \infty$  (due to Assumption 2.1), we have

$$\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) = \frac{\partial_\theta^\alpha r_{\theta, y_n}^0(x|\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda))}{\langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle} - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \partial_\theta^\beta p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) \langle S_{\theta, y_n}^{\alpha-\beta}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle.$$

Combining this with (124), (126), we get

$$\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) = s_{\theta, y_n}^\alpha(x|\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \partial_\theta^\beta p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) \langle S_{\theta, y_n}^{\alpha-\beta}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle. \tag{128}$$

Equation (128) can be interpreted as a recursion in  $|\alpha|$  which generates functions  $\{\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$ .<sup>7</sup> Equation (128) can also be considered as a particular case of (6) — to get (128), set  $\Lambda = \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)$ ,  $y = y_n$  in (6). Hence, comparing (128) with (6) and using (7), (9), (115), (116), we conclude

$$\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{0, n}(x|\lambda) = f_{\theta, y_n}^\alpha(x|\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)), \quad \tilde{P}_{\theta, \mathbf{y}}^{\alpha, n}(\lambda) = F_{\theta, y_n}^\alpha(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)), \quad \tilde{P}_{\theta, \mathbf{y}}^n(\lambda) = F_{\theta, y_n}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)). \tag{129}$$

---

<sup>7</sup>In (128),  $p_{\theta, \mathbf{y}}^{0, n}(x|\lambda)$  is the initial condition. At iteration  $k$  of recursion (128) ( $1 \leq k \leq p$ ), function  $\partial_\theta^\alpha p_{\theta, \mathbf{y}}^{0, n}(x|\lambda)$  is computed for multi-indices  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = k$  using the results obtained at the previous iterations.

Iterating (in  $n$ ) the last part of (129), we also get  $\tilde{P}_{\theta, \mathbf{y}}^n(\lambda) = F_{\theta, \mathbf{y}}^{0:n}(\mathcal{E}_\lambda)$ . Combining this with (11), (117), we deduce that the first two parts of (103) hold.

In the rest of the proof, we assume  $1 < |\alpha| \leq p$ . Owing to (101), (125), (127), we have

$$\partial_\theta^e \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) = \frac{\partial_\theta^e \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle}{\langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle} = \langle S_{\theta, y_n}^e(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle = \Psi_\theta^e(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)), \quad (130)$$

where  $e \in \mathbb{N}_0^d$ ,  $|e| = 1$ . Hence, we get

$$\partial_\theta^{e_\alpha} \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle = \partial_\theta^{e_\alpha} \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle$$

(as  $|e_\alpha| = 1$ ). Therefore, we have

$$\partial_\theta^\alpha \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle = \partial_\theta^{\alpha - e_\alpha} \left( \partial_\theta^{e_\alpha} \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle \right).$$

Consequently, Leibniz rule and (125), (127) imply

$$\begin{aligned} \partial_\theta^\alpha \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha - e_\alpha}} \binom{\alpha - e_\alpha}{\beta} \partial_\theta^{\beta + e_\alpha} \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \partial_\theta^{\alpha - \beta - e_\alpha} \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle \\ &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} \partial_\theta^\beta \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \partial_\theta^{\alpha - \beta} \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle \\ &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} \partial_\theta^\beta \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \langle S_{\theta, y_n}^{\alpha - \beta}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle. \end{aligned}$$

As  $\langle S_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle = 1$  (due to (5), (8)), the same arguments then yield

$$\begin{aligned} \partial_\theta^\alpha \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) &= - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} \partial_\theta^\beta \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \langle S_{\theta, y_n}^{\alpha - \beta}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle \\ &\quad + \frac{\partial_\theta^\alpha \langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle}{\langle R_{\theta, y_n}^0(\tilde{P}_{\theta, \mathbf{y}}^{0, n-1}(\lambda)) \rangle} \\ &= - \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} \partial_\theta^\beta \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \langle S_{\theta, y_n}^{\alpha - \beta}(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle \\ &\quad + \langle S_{\theta, y_n}^\alpha(\tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) \rangle. \end{aligned} \quad (131)$$

Equation (131) can be viewed as a recursion in  $|\alpha|$  which generates functions  $\{\partial_\theta^\alpha \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) : \alpha \in \mathbb{N}_0^d, 1 < |\alpha| \leq p\}$ .<sup>8</sup> Equation (131) can also be considered as a special case of (102) — to get (131), set  $\Lambda = \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)$ ,  $y = y_n$  in (102). Hence, comparing (131) with (101), (102), we conclude

$$\partial_\theta^\alpha \Psi_\theta^0(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)) = \Psi_\theta^\alpha(y_n, \tilde{P}_{\theta, \mathbf{y}}^{n-1}(\lambda)). \quad (132)$$

Using (130), (132), we deduce that the last part of (103) holds.  $\square$

**Proof of Theorem 2.1.** Let  $m \geq 0$  be any (fixed) integer, while  $\mathbf{y} = \{y_n\}_{n \geq 1}$ ,  $\mathbf{y}' = \{y'_n\}_{n \geq 1}$  are any sequences in  $\mathcal{Y}$  satisfying  $y'_n = y_{n+m}$  for  $n > m$ . Then, using (3), it is straightforward to verify  $p_{\theta, \mathbf{y}}^{m:n}(x|\lambda) = p_{\theta, \mathbf{y}'}^{0:n-m}(x|\lambda)$  for  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $\lambda \in \mathcal{P}(\mathcal{X})$ ,  $n > m$ . Consequently, Proposition 7.1 implies that (14) holds for the same  $\theta$ ,  $x$ ,  $\lambda$ ,  $n$ ,  $m$  and  $B \in \mathcal{B}(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .  $\square$

<sup>8</sup>In (131), functions  $\{\partial_\theta^\alpha \Psi_\theta^0(y_{n+1}, \tilde{P}_{\theta, \mathbf{y}}^n(\lambda)) : \alpha \in \mathbb{N}_0^d, |\alpha| = 1\}$  are the initial conditions. At iteration  $k$  of (131) ( $1 < k \leq p$ ), function  $\partial_\theta^\alpha \Psi_\theta^0(y_{n+1}, \tilde{P}_{\theta, \mathbf{y}}^n(\lambda))$  is computed for multi-indices  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = k$  using the results obtained at the previous iterations.

**Lemma 7.1.** *Let Assumptions 2.1 and 3.1 hold. Then, there exists a real number  $C_8 \in [1, \infty)$  (depending only on  $\varepsilon$ ) such that*

$$|\Psi_\theta^0(y, \Lambda)| \leq C_8 \varphi(y), \quad |\Psi_\theta^0(y, \Lambda) - \Psi_\theta^0(y, \Lambda')| \leq C_8 \varphi(y) \|\Lambda - \Lambda'\|$$

for all  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ .

*Proof.* Throughout the proof, the following notation is used.  $\tilde{C}$  is the real number defined by  $\tilde{C} = 1 + |\log \varepsilon|$ , while  $C_8$  is the real number defined by  $C_8 = 2C_1\tilde{C}$  ( $\varepsilon$ ,  $C_1$  are specified in Assumption 2.1 and Lemma 5.1).  $\theta$ ,  $y$  are any elements of  $\Theta$ ,  $\mathcal{Y}$  (respectively).  $\lambda, \lambda'$  are any elements of  $\mathcal{P}(\mathcal{X})$ , while  $\Lambda = \{\lambda_\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$ ,  $\Lambda' = \{\lambda'_\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$  are any elements of  $\mathcal{L}_0(\mathcal{X})$ .

Relying on Assumption 2.1, we conclude

$$\varepsilon \mu_\theta(\mathcal{X}|y) \leq \int \left( \int r_\theta(y, x'|x) \mu(dx') \right) \lambda(dx) \leq \frac{\mu_\theta(\mathcal{X}|y)}{\varepsilon}.$$

Consequently, Assumption 3.1 and (8) imply

$$|\log(\langle R_{\theta,y}^0(\lambda) \rangle)| \leq |\log \varepsilon| + |\log \mu_\theta(\mathcal{X}|y)| \leq \tilde{C} + \varphi(y) \leq 2\tilde{C}\varphi(y).$$

Therefore, (101) yields

$$|\Psi_\theta^0(y, \Lambda)| = |\log(\langle R_{\theta,y}^0(\lambda_0) \rangle)| \leq 2\tilde{C}\varphi(y) \leq C_8\varphi(y).$$

Moreover, using Lemma 5.1, we deduce

$$\left| \frac{\langle R_{\theta,y}^0(\lambda) \rangle}{\langle R_{\theta,y}^0(\lambda') \rangle} - 1 \right| = \left| \frac{\langle R_{\theta,y}^0(\lambda - \lambda') \rangle}{\langle R_{\theta,y}^0(\lambda') \rangle} \right| \leq \frac{\|\langle R_{\theta,y}^0(\lambda - \lambda') \rangle\|}{\langle R_{\theta,y}^0(\lambda') \rangle} \leq C_1 \|\lambda - \lambda'\|.$$

Consequently, we have

$$\log \left( \frac{\langle R_{\theta,y}^0(\lambda) \rangle}{\langle R_{\theta,y}^0(\lambda') \rangle} \right) \leq \left| \frac{\langle R_{\theta,y}^0(\lambda) \rangle}{\langle R_{\theta,y}^0(\lambda') \rangle} - 1 \right| \leq C_1 \|\lambda - \lambda'\|. \quad (133)$$

Reverting the roles of  $\lambda$ ,  $\lambda'$ , we get

$$-\log \left( \frac{\langle R_{\theta,y}^0(\lambda) \rangle}{\langle R_{\theta,y}^0(\lambda') \rangle} \right) = \log \left( \frac{\langle R_{\theta,y}^0(\lambda') \rangle}{\langle R_{\theta,y}^0(\lambda) \rangle} \right) \leq \left| \frac{\langle R_{\theta,y}^0(\lambda') \rangle}{\langle R_{\theta,y}^0(\lambda) \rangle} - 1 \right| \leq C_1 \|\lambda - \lambda'\|. \quad (134)$$

Owing to (133), (134), we have

$$\left| \log \left( \frac{\langle R_{\theta,y}^0(\lambda) \rangle}{\langle R_{\theta,y}^0(\lambda') \rangle} \right) \right| \leq C_1 \|\lambda - \lambda'\|.$$

Hence, we get

$$|\Psi_\theta^0(y, \Lambda) - \Psi_\theta^0(y, \Lambda')| = \left| \log \left( \frac{\langle R_{\theta,y}^0(\lambda_0) \rangle}{\langle R_{\theta,y}^0(\lambda'_0) \rangle} \right) \right| \leq C_1 \|\lambda_0 - \lambda'_0\| \leq C_8 \varphi(y) \|\Lambda - \Lambda'\|$$

(as  $\varphi(y) \geq 1$ ,  $\|\lambda_0 - \lambda'_0\| \leq \|\Lambda - \Lambda'\|$ ). □

**Lemma 7.2.** *Let Assumptions 2.1 and 2.2 hold. Then, there exists a real number  $C_9 \in [1, \infty)$  (depending only on  $\varepsilon$ ,  $p$ ) such that*

$$|\Psi_\theta^\alpha(y, \Lambda)| \leq C_9 (\psi(y) \|\Lambda\|)^p, \quad |\Psi_\theta^\alpha(y, \Lambda) - \Psi_\theta^\alpha(y, \Lambda')| \leq C_9 \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^p$$

for all  $\theta \in \Theta$ ,  $y \in \mathcal{Y}$ ,  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$  and any multi-index  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ ,  $|\alpha| \leq p$ .

*Proof.* Throughout the proof, the following notation is used.  $\theta, y$  are any elements of  $\Theta, \mathcal{Y}$  (respectively).  $\tilde{C}_1, \tilde{C}_2$  are the real numbers defined by  $\tilde{C}_1 = 2^p C_2, \tilde{C}_2 = 3\tilde{C}_1^2$ , while  $C_9$  is the real number defined by  $C_9 = \exp(\tilde{C}_2 p)$  ( $C_2$  is specified in Lemma 5.2).  $B_\alpha$  is the real number defined by  $B_\alpha = \exp(\tilde{C}_2 |\alpha|)$  for  $\alpha \in \mathbb{N}_0^d$ .

Let  $\gamma$  be any element of  $\mathbb{N}_0^d \setminus \{\mathbf{0}\}$  satisfying  $|\gamma| \leq p$ . Then, it easy to conclude  $B_\gamma \leq \exp(\tilde{C}_2) \leq 3\tilde{C}_1^2$ . Consequently, Lemma 5.2 and (8) imply

$$\begin{aligned} \left| \langle S_{\theta,y}^\gamma(\Lambda) \rangle - \langle S_{\theta,y}^\gamma(\Lambda') \rangle \right| &\leq C_2 \sum_{\substack{\delta \in \mathbb{N}_0^d \\ \delta \leq \gamma}} (\psi(y))^{|\gamma-\delta|} (\|\lambda_\delta - \lambda'_\delta\| + \|\lambda_0 - \lambda'_0\| \|\lambda'_\delta\|) \\ &\leq 2^{|\gamma|} C_2 (\psi(y))^{|\gamma|} \|\Lambda - \Lambda'\| (\|\Lambda\| + \|\Lambda'\|) \\ &\leq \tilde{C}_1 \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\gamma|} \\ &\leq \frac{B_\gamma \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\gamma|}}{3\tilde{C}_1} \end{aligned} \quad (135)$$

for  $\Lambda = \{\lambda_\delta : \delta \in \mathbb{N}_0^d, |\delta| \leq p\} \in \mathcal{L}_0(\mathcal{X})$ ,  $\Lambda' = \{\lambda'_\delta : \delta \in \mathbb{N}_0^d, |\delta| \leq p\} \in \mathcal{L}_0(\mathcal{X})$  (as  $\|\Lambda\| \geq 1, \|\Lambda'\| \geq \|\lambda'_\delta\|, \|\Lambda - \Lambda'\| \geq \|\lambda_\delta - \lambda'_\delta\|$ ). The same arguments yield

$$\left| \langle S_{\theta,y}^\gamma(\Lambda) \rangle \right| \leq C_2 \sum_{\substack{\delta \in \mathbb{N}_0^d \\ \delta \leq \gamma}} (\psi(y))^{|\gamma-\delta|} \|\lambda_\delta\| \leq 2^{|\gamma|} C_2 (\psi(y))^{|\gamma|} \|\Lambda\| \leq \tilde{C}_1 (\psi(y) \|\Lambda\|)^{|\gamma|} \leq \frac{B_\alpha (\psi(y) \|\Lambda\|)^{|\gamma|}}{3\tilde{C}_1} \quad (136)$$

for the same  $\Lambda$ .

To prove the lemma, it is sufficient to show

$$|\Psi_\theta^\alpha(y, \Lambda)| \leq B_\alpha (\psi(y) \|\Lambda\|)^{|\alpha|}, \quad |\Psi_\theta^\alpha(y, \Lambda) - \Psi_\theta^\alpha(y, \Lambda')| \leq B_\alpha \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\alpha|} \quad (137)$$

for  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ ,  $|\alpha| \leq p$ . We prove (137) by mathematical induction in  $|\alpha|$ . When  $|\alpha| = 1$ , (135), (136) imply that (137) is true for all  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ . Now, the induction hypothesis is formulated: Suppose that (137) holds for some  $l \in \mathbb{N}_0^d$  and all  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$  satisfying  $1 \leq l < p$ ,  $|\alpha| \leq l$ . Then, to prove (137), it is sufficient to show (137) for all  $\Lambda, \Lambda' \in \mathcal{L}_0(\mathcal{X})$ ,  $\alpha \in \mathbb{N}_0^d$  satisfying  $|\alpha| = l + 1$ . In what follows in the proof,  $\Lambda, \Lambda'$  are any elements of  $\mathcal{L}_0(\mathcal{X})$ .  $\alpha$  is any element of  $\mathbb{N}_0^d$  satisfying  $|\alpha| = l + 1$ , while  $\beta$  is any element of  $\mathbb{N}_0^d \setminus \{\mathbf{0}, \alpha\}$  fulfilling  $\beta \leq \alpha$ .

Since  $\beta \leq \alpha$ ,  $\beta \neq \mathbf{0}$ ,  $\beta \neq \alpha$ , we have  $1 \leq |\beta| \leq |\alpha| - 1 = l$ . Then, owing to the induction hypothesis, we have

$$\max \left\{ \frac{|\Psi_\theta^\beta(y, \Lambda)|}{(\psi(y) \|\Lambda\|)^{|\beta|}}, \frac{|\Psi_\theta^\beta(y, \Lambda')|}{(\psi(y) \|\Lambda'\|)^{|\beta|}} \right\} \leq B_\beta \leq \frac{B_\alpha}{3\tilde{C}_1^2}, \quad (138)$$

$$\frac{|\Psi_\theta^\beta(y, \Lambda) - \Psi_\theta^\beta(y, \Lambda')|}{(\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\beta|}} \leq B_\beta \|\Lambda - \Lambda'\| \leq \frac{B_\alpha \|\Lambda - \Lambda'\|}{3\tilde{C}_1^2}. \quad (139)$$

Consequently, (136) implies

$$|\Psi_\theta^\beta(y, \Lambda)| |\langle S_{\theta,y}^{\alpha-\beta}(\Lambda) \rangle| \leq \frac{B_\beta B_{\alpha-\beta} (\psi(y) \|\Lambda\|)^{|\alpha|}}{3\tilde{C}_1} \leq \frac{B_\alpha (\psi(y) \|\Lambda\|)^{|\alpha|}}{3\tilde{C}_1} \quad (140)$$

(as  $|\alpha| = |\beta| + |\alpha - \beta|$ ). Similarly, (135), (136), (138), (139) yield

$$|\Psi_\theta^\beta(y, \Lambda) - \Psi_\theta^\beta(y, \Lambda')| |\langle S_{\theta,y}^{\alpha-\beta}(\Lambda') \rangle| \leq \frac{B_\alpha \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\alpha|}}{3\tilde{C}_1}, \quad (141)$$

$$|\Psi_\theta^\beta(y, \Lambda)| |\langle S_{\theta,y}^{\alpha-\beta}(\Lambda) \rangle - \langle S_{\theta,y}^{\alpha-\beta}(\Lambda') \rangle| \leq \frac{B_\alpha \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\alpha|}}{3\tilde{C}_1}. \quad (142)$$

Using (102), (136), (140), we conclude

$$\begin{aligned}
|\Psi_\theta^\alpha(y, \Lambda)| &\leq |\langle S_{\theta, y}^\alpha(\Lambda) \rangle| + \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} |\Psi_\theta^\beta(y, \Lambda)| |\langle S_{\theta, y}^{\alpha - \beta}(\Lambda) \rangle| \\
&\leq \frac{2^{|\alpha|} B_\alpha(\psi(y) \|\Lambda\|)^{|\alpha|}}{\tilde{C}_1} \\
&\leq B_\alpha(\psi(y) \|\Lambda\|)^{|\alpha|}.
\end{aligned}$$

(as  $\tilde{C}_1 \geq 2^{|\alpha|}$ ). Relying on (102), (135), (141), (142), we deduce

$$\begin{aligned}
|\Psi_\theta^\alpha(y, \Lambda) - \Psi_\theta^\alpha(y, \Lambda')| &\leq \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} |\Psi_\theta^\beta(y, \Lambda)| |\langle S_{\theta, y}^{\alpha - \beta}(\Lambda) \rangle - \langle S_{\theta, y}^{\alpha - \beta}(\Lambda') \rangle| \\
&\quad + \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{\alpha\} \\ e_\alpha \leq \beta \leq \alpha}} \binom{\alpha - e_\alpha}{\beta - e_\alpha} |\Psi_\theta^\beta(y, \Lambda) - \Psi_\theta^\beta(y, \Lambda')| |\langle S_{\theta, y}^{\alpha - \beta}(\Lambda') \rangle| \\
&\quad + |\langle S_{\theta, y}^\alpha(\Lambda) \rangle - \langle S_{\theta, y}^\alpha(\Lambda') \rangle| \\
&\leq \frac{2^{|\alpha|} B_\alpha \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\alpha|}}{\tilde{C}_1} \\
&\leq B_\alpha \|\Lambda - \Lambda'\| (\psi(y) (\|\Lambda\| + \|\Lambda'\|))^{|\alpha|}.
\end{aligned}$$

Hence, (137) holds for  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = l+1$ . Then, the lemma directly follows by the principle of mathematical induction.  $\square$

**Proof of Theorem 3.1.** Let  $w = p(p+1)$ . Using Theorem 2.3 and Lemmas 7.1, 7.2, we conclude that for each multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ , there exists a function  $\psi_\theta^\alpha$  which maps  $\theta$  to  $\mathbb{R}$  and satisfies

$$\psi_\theta^\alpha = \lim_{n \rightarrow \infty} (\tilde{\Pi}^n \Psi^\alpha)_\theta(x, y, \Lambda) \quad (143)$$

for  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $\Lambda \in \mathcal{L}_0(\mathcal{X})$ . Relying on the same arguments, we deduce that there also exist real numbers  $\rho \in (0, 1)$ ,  $\tilde{C}_1 \in [1, \infty)$  (depending only on  $\varepsilon$ ,  $\delta$ ,  $K_0$ ,  $M_0$ ) such that

$$\left| (\tilde{\Pi}^n \Psi^\alpha)_\theta(x, y, \Lambda) - \psi_\theta^\alpha \right| \leq \tilde{C}_1 \rho^n \psi^u(y) \|\Lambda\|^w \quad (144)$$

for the same  $\theta$ ,  $x$ ,  $y$ ,  $\Lambda$  and  $n \geq 1$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$  ( $u$  is specified in Assumption 3.1).

Throughout the rest of the proof, the following notation is used.  $\theta$  is any element of  $\Theta$ , while  $x$ ,  $y$ ,  $\lambda$  are any elements of  $\mathcal{X}$ ,  $\mathcal{Y}$ ,  $\mathcal{P}(\mathcal{X})$  (respectively).  $\alpha$  is any element of  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .  $n$  is any (strictly) positive integer.

Owing to Assumption 3.1, we have

$$\max \{E(\varphi(Y_n)), E(\psi^u(Y_n))\} \leq E(\varphi(Y_n) \psi^u(Y_n)) = E\left(\int \varphi(y) \psi^u(y) Q(X_n, dy)\right) \leq M_0. \quad (145)$$

Due to the same assumption, we also have

$$\begin{aligned}
E(\psi^p(Y_k) \psi^u(Y_l) | X_1 = x, Y_1 = y) &= \psi^u(y) E\left(\int \psi^p(y) Q(X_k, dy)\right) \leq M_0 \psi^u(y), \\
E(\psi^p(Y_l) \psi^u(Y_k) | X_1 = x, Y_1 = y) &= E\left(\int \psi^p(y) Q(X_l, dy) \int \psi^u(y) Q(X_k, dy)\right) \leq M_0^2
\end{aligned}$$

for  $l > k > 1$ . Therefore, we get

$$E\left(\psi^p(Y_{n+1}) \sum_{k=1}^n \psi^u(Y_k) \middle| X_1 = x, Y_1 = y\right) \leq M_0^2 n + M_0 \psi^u(y) < \infty. \quad (146)$$

Using (3), (10), (18), (101), (102), it is straightforward to verify

$$\begin{aligned}\log q_\theta^n(Y_{1:n}|\lambda) &= \sum_{k=1}^{n-1} \log \left( \int \int r_\theta(Y_{k+1}, x''|x') p_{\theta, \mathbf{Y}}^{0:k}(x'|\lambda) \mu(dx'') \mu(dx') \right) \\ &\quad + \log \left( \int \int r_\theta(Y_1, x'|x) \mu(dx') \lambda(dx) \right) \\ &= \sum_{k=0}^{n-1} \Psi_\theta^0(Y_{k+1}, F_{\theta, \mathbf{Y}}^{0:k}(\mathcal{E}_\lambda))\end{aligned}$$

(here,  $\mathbf{Y}$  denotes stochastic process  $\{Y_n\}_{n \geq 1}$ , i.e.,  $\mathbf{Y} = \{Y_n\}_{n \geq 1}$ ). It is also easy to show

$$(\tilde{\Pi}^n \Psi^\alpha)_\theta(x, y, \mathcal{E}_\lambda) = E \left( \Psi_\theta^\alpha(Y_{n+1}, F_{\theta, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda)) \mid X_1 = x, Y_1 = y \right).$$

Therefore, we have

$$E \left( \log q_\theta^n(Y_{1:n}|\lambda) \mid X_1 = x, Y_1 = y \right) = \sum_{k=1}^{n-1} (\tilde{\Pi}^k \Psi^0)_\theta(x, y, \mathcal{E}_\lambda) + \Psi_\theta^0(y, \mathcal{E}_\lambda). \quad (147)$$

Consequently, Lemma 7.1 and (144) imply

$$\begin{aligned}\left| E \left( \frac{1}{n} \log q_\theta^n(Y_{1:n}|\lambda) \mid X_1 = x, Y_1 = y \right) - \psi_\theta^0 \right| &\leq \frac{1}{n} \sum_{k=1}^{n-1} \left| (\tilde{\Pi}^k \Psi^0)_\theta(x, y, \mathcal{E}_\lambda) - \psi_\theta^0 \right| + \frac{|\psi_\theta^0| + |\Psi_\theta^0(y, \mathcal{E}_\lambda)|}{n} \\ &\leq \frac{\tilde{C}_1 \psi^u(y)}{n} \sum_{k=1}^{n-1} \rho^k + \frac{|\psi_\theta^0| + C_8 \phi(y)}{n} \\ &\leq \frac{\tilde{C}_1 \psi^u(y)}{n(1-\rho)} + \frac{|\psi_\theta^0| + C_8 \phi(y)}{n}.\end{aligned}$$

Then, (145) yields

$$\begin{aligned}\left| E \left( \frac{1}{n} \log q_\theta^n(Y_{1:n}|\lambda) \right) - \psi_\theta^0 \right| &\leq E \left( \left| E \left( \frac{1}{n} \log q_\theta^n(Y_{1:n}|\lambda) \mid X_1, Y_1 \right) - \psi_\theta^0 \right| \right) \\ &\leq \frac{\tilde{C}_1 E(\psi^u(Y_1))}{n(1-\rho)} + \frac{|\psi_\theta^0| + C_8 E(\phi(Y_1))}{n} \\ &\leq \frac{\tilde{C}_1 M_0}{n(1-\rho)} + \frac{|\psi_\theta^0| + C_8 M_0}{n}.\end{aligned}$$

Therefore, we get

$$\lim_{n \rightarrow \infty} E \left( \frac{1}{n} \log q_\theta^n(Y_{1:n}|\lambda) \right) = \psi_\theta^0. \quad (148)$$

Let  $\tilde{C}_2 = \max\{A_\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq p\}$  ( $A_\alpha$  is specified in Proposition 5.3). Owing to Proposition 5.3 and Lemma 7.2, we have

$$\begin{aligned}|\Psi_\theta^\alpha(Y_{n+1}, F_{\theta, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda))| &\leq C_9 \psi^p(Y_{n+1}) \|F_{\theta, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda)\|^p \leq \tilde{C}_2^p C_9 \psi^p(Y_{n+1}) (\Psi_\mathbf{Y}^{0:n})^u \\ &\leq \tilde{C}_2^p C_9 n^u \psi^p(Y_{n+1}) \sum_{k=1}^n \psi^u(Y_k)\end{aligned}$$

(as  $\Psi_\mathbf{Y}^{0:n} \geq 1$ ,  $u > p^2$ ). Consequently, Proposition 7.1, Lemma A3.1 and (146), (147) imply that  $(\tilde{\Pi}^n \Psi^0)_\theta(x, y, \mathcal{E}_\lambda)$  is  $p$ -times differentiable in  $\theta$  and satisfies

$$\begin{aligned}\partial_\theta^\alpha (\tilde{\Pi}^n \Psi^0)_\theta(x, y, \mathcal{E}_\lambda) &= E \left( \partial_\theta^\alpha \Psi_\theta^0(Y_{n+1}, F_{\theta, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda)) \mid X_1 = x, Y_1 = y \right) \\ &= E \left( \Psi_\theta^\alpha(Y_{n+1}, F_{\theta, \mathbf{Y}}^{0:n}(\mathcal{E}_\lambda)) \mid X_1 = x, Y_1 = y \right) \\ &= (\tilde{\Pi}^n \Psi^\alpha)_\theta(x, y, \mathcal{E}_\lambda).\end{aligned}$$



Then, the uniform convergence theorem and (144) yield that  $\psi_\theta^0$  is  $p$ -times differentiable in  $\theta$  and satisfies  $\partial_\theta^\alpha \psi_\theta^0 = \psi_\theta^\alpha$ . Combining this with (148), we conclude that there exists function  $l(\theta)$  with the properties specified in the statement of the theorem.  $\square$

## 8. Proof of Corollaries 4.1 and 4.2

Throughout this section, we rely on the following notation.  $\tilde{A}'_\theta(x'|x)$ ,  $\tilde{B}'_\theta(x)$ ,  $\tilde{B}_\theta(x)$ ,  $\tilde{C}'_\theta(y|x)$ ,  $\tilde{D}'_\theta(x)$  and  $\tilde{D}_\theta(x)$  are the functions defined by

$$\begin{aligned}\tilde{A}'_\theta(x'|x) &= x' - A_\theta(x), & \tilde{B}'_\theta(x) &= \text{adj} B_\theta(x), & \tilde{B}_\theta(x) &= \det B_\theta(x), \\ \tilde{C}'_\theta(y|x) &= y - C_\theta(x), & \tilde{D}'_\theta(x) &= \text{adj} D_\theta(x), & \tilde{D}_\theta(x) &= \det D_\theta(x)\end{aligned}$$

for  $\theta \in \tilde{\Theta}$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ .  $\tilde{A}_\theta(x'|x)$ ,  $\tilde{C}_\theta(y|x)$ ,  $U_\theta(x'|x)$  and  $V_\theta(y|x)$  are the functions defined by

$$\begin{aligned}\tilde{A}_\theta(x'|x) &= \tilde{B}'_\theta(x) \tilde{A}'_\theta(x'|x), & U_\theta(x'|x) &= \frac{\tilde{A}_\theta(x'|x)}{\tilde{B}_\theta(x)}, \\ \tilde{C}_\theta(y|x) &= \tilde{D}'_\theta(x) \tilde{C}'_\theta(y|x), & V_\theta(y|x) &= \frac{\tilde{C}_\theta(y|x)}{\tilde{D}_\theta(x)}.\end{aligned}$$

$u_\theta(x'|x)$ ,  $\bar{u}_\theta(x)$ ,  $v_\theta(y|x)$  and  $\bar{v}_\theta(x)$  are the functions defined by

$$\begin{aligned}u_\theta(x'|x) &= r(U_\theta(x'|x)), & \bar{u}_\theta(x) &= \int_{\mathcal{X}} u_\theta(x''|x) dx'', \\ v_\theta(y|x) &= s(V_\theta(y|x)), & \bar{v}_\theta(x) &= \int_{\mathcal{Y}} v_\theta(y'|x) dy'.\end{aligned}$$

Then, it is easy to show

$$U_\theta(x'|x) = B_\theta^{-1}(x)(x' - A_\theta(x)), \quad V_\theta(y|x) = D_\theta^{-1}(x)(y - C_\theta(x))$$

for all  $\theta \in \tilde{\Theta}$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ . It is also easy to demonstrate

$$p_\theta(x'|x) = \frac{u_\theta(x'|x)}{\bar{u}_\theta(x)}, \quad q_\theta(y|x) = \frac{v_\theta(y|x)}{\bar{v}_\theta(x)}.$$

**Lemma 8.1.** *Let Assumptions 4.1 – 4.4 hold. Then,  $p_\theta(x'|x)$  and  $q_\theta(y|x)$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ . Moreover, there exist real numbers  $\varepsilon_1 \in (0, 1)$ ,  $K_1 \in [1, \infty)$  such that*

$$\min \{p_\theta(x'|x), q_\theta(y|x)\} \geq \varepsilon_1, \quad \max \{|\partial_\theta^\alpha p_\theta(x'|x)|, |\partial_\theta^\alpha q_\theta(y|x)|\} \leq K_1 \quad (149)$$

for all  $\theta \in \Theta$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .

*Proof.* Throughout the proof,  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ . It is easy to notice that  $\tilde{B}_\theta(x)$  and the entries of  $\tilde{B}'_\theta(x)$  are polynomial in the entries of  $B_\theta(x)$ . It is also easy to notice that  $\tilde{D}_\theta(x)$  and the entries of  $\tilde{D}'_\theta(x)$  are polynomial in the entries of  $D_\theta(x)$ . Consequently, Assumptions 4.2, 4.3 imply that  $\partial_\theta^\alpha \tilde{A}_\theta(x'|x)$ ,  $\partial_\theta^\alpha \tilde{B}_\theta(x)$  exist and are continuous in  $(\theta, x, x')$ ,  $(\theta, x)$  on  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{X}$ ,  $\tilde{\Theta} \times \mathcal{X}$ . The same assumptions also imply that  $\partial_\theta^\alpha \tilde{C}_\theta(y|x)$ ,  $\partial_\theta^\alpha \tilde{D}_\theta(x)$  exist and are continuous in  $(\theta, x, y)$ ,  $(\theta, x)$  on  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{Y}$ ,  $\tilde{\Theta} \times \mathcal{X}$ . As  $\tilde{B}_\theta(x)$ ,  $\tilde{D}_\theta(x)$  are non-zero (due to Assumption 4.1), we conclude from Lemma A2.1 (see Appendix 2) that  $\partial_\theta^\alpha \tilde{U}_\theta(x'|x)$ ,  $\partial_\theta^\alpha \tilde{V}_\theta(y|x)$  exist and are continuous in  $(\theta, x, x')$ ,  $(\theta, x, y)$  on  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{X}$ ,  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{Y}$ . Then, using Assumption 4.2 and Lemma A1.1 (see Appendix 1), we deduce that  $\partial_\theta^\alpha \tilde{u}_\theta(x'|x)$ ,  $\partial_\theta^\alpha \tilde{v}_\theta(y|x)$  exist and are continuous in  $(\theta, x, x')$ ,  $(\theta, x, y)$  on  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{X}$ ,  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{Y}$ .

Let  $\theta$  be any element of  $\Theta$ . Moreover, let  $x, x'$  be any elements of  $\mathcal{X}$ , while  $y$  is any element of  $\mathcal{Y}$ . Since  $\Theta$  is bounded and  $\text{cl}\Theta \subset \tilde{\Theta}$ , Assumptions 4.1, 4.4 imply that there exist real numbers  $\delta \in (0, 1)$ ,  $\tilde{C} \in [1, \infty)$  (independent of  $\theta, x, x', y, \alpha$ ) such that

$$\min \{u_\theta(x'|x), v_\theta(y|x)\} \geq \delta, \quad \max \{|\partial_\theta^\alpha u_\theta(x'|x)|, |\partial_\theta^\alpha v_\theta(y|x)|\} \leq \tilde{C}. \quad (150)$$

Consequently, Lemma A3.1 (see Appendix 3) yields that  $\partial_\theta^\alpha \bar{u}_\theta(x)$ ,  $\partial_\theta^\alpha \bar{v}_\theta(x)$  exist. Moreover, combining Assumption 4.4 and (150), we get

$$\bar{u}_\theta(x) = \int_{\mathcal{X}} u_\theta(x'|x) dx' \geq \delta m(\mathcal{X}) > 0, \quad \bar{v}_\theta(x) = \int_{\mathcal{Y}} v_\theta(y|x) dy \geq \delta m(\mathcal{Y}) > 0, \quad (151)$$

where  $m(\mathcal{X})$ ,  $m(\mathcal{Y})$  are the Lebesgue measures of  $\mathcal{X}$ ,  $\mathcal{Y}$  (respectively). Then, using Lemma A2.1, we conclude that  $\partial_\theta^\alpha p_\theta(x'|x)$ ,  $\partial_\theta^\alpha q_\theta(y|x)$  exist. Relying on the same lemma and (150), (151), we deduce that there exists a real number  $K_1 \in [1, \infty)$  with the properties specified in the lemma's statement.  $\square$

**Proof of Corollary 4.1.** Throughout the proof, the following notation is used.  $\varepsilon$ ,  $\tilde{C}_1$ ,  $\tilde{C}_2$ ,  $\tilde{C}_3$  are the real numbers defined by  $\varepsilon = \min\{\varepsilon_1^2, K_1^{-2}\}$ ,  $\tilde{C}_1 = 2K_1^2\varepsilon_1^{-2}$ ,  $\tilde{C}_2 = K_1^2$ ,  $\tilde{C}_3 = 1 + |\log \mu(\mathcal{X})|$  ( $\varepsilon_1$ ,  $K_1$  are specified in Lemma 8.1).  $r$ ,  $u$ ,  $v$  are the real numbers specified in Assumptions 2.6, 3.2.  $\psi(y)$ ,  $\phi(x, y)$ ,  $\varphi(y)$  and  $\mu_\theta(dx|y)$  are the functions and the measure defined by

$$\psi(y) = \tilde{C}_1, \quad \phi(x, y) = \tilde{C}_2, \quad \varphi(y) = \tilde{C}_3, \quad \mu_\theta(B|y) = \mu(B)$$

for  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $B \in \mathcal{B}(\mathcal{X})$  ( $\mu(dx)$  is specified in Subsection 2.1).  $r_\theta(y, x'|x)$  has the same meaning as in (1), while  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  are defined in (20).  $\theta$  is any element of  $\Theta$ , while  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .  $x, x'$  are any elements of  $\mathcal{X}$ , while  $y$  is any element of  $\mathcal{Y}$ .

(i) Owing to Lemma 8.1, we have

$$\varepsilon_1^2 \leq r_\theta(y, x'|x) \leq K_1^2. \quad (152)$$

Consequently, we get

$$\int_B r_\theta(y, x'|x) \mu(dx') \geq \varepsilon_1^2 \mu(B) \geq \varepsilon \mu_\theta(B|y), \quad \int_B r_\theta(y, x'|x) \mu(dx') \leq K_1^2 \mu(B) \leq \frac{1}{\varepsilon} \mu_\theta(B|y)$$

for  $B \in \mathcal{B}(\mathcal{X})$ . We also get

$$r_\theta(y, x'|x) \leq \tilde{C}_2 = \phi(y, x'), \quad \int \phi(y, x) \mu(dx) = \tilde{C}_2 \mu(\mathcal{X}) < \infty.$$

Hence, Assumptions 2.1, 2.3 hold for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20).

Due to Leibniz formula and Lemma 8.1, we have

$$|\partial_\theta^\alpha r_\theta(y, x'|x)| \leq \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \left| \partial_\theta^\beta q_\theta(y|x') \right| \left| \partial_\theta^{\alpha-\beta} p_\theta(x'|x) \right| \leq K_1^2 \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} = 2^{|\alpha|} K_1^2.$$

Then, (152) implies

$$|\partial_\theta^\alpha r_\theta(y, x'|x)| \leq 2^{|\alpha|} K_1^2 \varepsilon_1^{-2} r_\theta(y, x'|x) \leq (\psi(y))^{|\alpha|} r_\theta(y, x'|x).$$

Thus, Assumption 2.2 holds for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20). Consequently, all conclusions of Theorems 2.1, 2.2 are true for the model introduced in Section 4.

(ii) Owing to (21), we have

$$\int \varphi(x, y) \psi^r(y) Q(x, dy) \leq \tilde{C}_1^r \sup_{x' \in \mathcal{X}} \int \varphi(x', y) Q(x', dy) < \infty.$$

Hence, in addition to Assumptions 2.1 – 2.3, Assumptions 2.4 – 2.6 also hold for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20). Therefore, all conclusions of Theorem 2.3 are true for the model introduced in Section 4.

(iii) It is easy to conclude

$$|\log \mu_\theta(\mathcal{X}|y)| = |\log \mu(\mathcal{X})| \leq \tilde{C}_3 = \varphi(y).$$

It is also easy to deduce

$$\int \varphi(y)\psi^u(y)Q(x, dy) = \tilde{C}_1^u \tilde{C}_3 < \infty, \quad \int \psi^v(y)Q(x, dy) = \tilde{C}_1^v < \infty.$$

Thus, in addition to Assumptions 2.1 – 2.3, Assumptions 2.4, 3.1, 3.2 also hold for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20). Consequently, all conclusions of Theorem 3.1 are true for the model introduced in Section 4.  $\square$

**Lemma 8.2.** (i) Let Assumptions 4.1 – 4.3 and 4.5 hold. Then,  $p_\theta(x'|x)$  and  $q_\theta(y|x)$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ . Moreover, there exist real numbers  $\varepsilon_2 \in (0, 1)$ ,  $K_2, K_3 \in [1, \infty)$  such that

$$p_\theta(x'|x) \geq \varepsilon_2, \quad |\partial_\theta^\alpha p_\theta(x'|x)| \leq K_2, \quad q_\theta(y|x) \leq K_3, \quad |\partial_\theta^\alpha q_\theta(y|x)| \leq K_3 q_\theta(y|x)(1 + \|y\|)^{2|\alpha|} \quad (153)$$

for all  $\theta \in \Theta$ ,  $x, x' \in \mathcal{X}$ ,  $y \in \mathcal{Y}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .

(ii) Let Assumptions 4.1 – 4.3, 4.5 and 4.6 hold. Then, there exist a real number  $K_4 \in [1, \infty)$  such that

$$|\log q_\theta(y|x)| \leq K_4(1 + \|y\|)^2 \quad (154)$$

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

*Proof.* Throughout the proof, the following notation is used.  $\theta$  is any element of  $\Theta$ , while  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .  $x, x'$  are any elements of  $\mathcal{X}$ , while  $y$  is any element of  $\mathcal{Y}$ .

(i) Using the same arguments as in the proof of Lemma 8.1, it can be shown that  $\partial_\theta^\alpha p_\theta(x'|x)$  exists. Relying on the same arguments, it can also be demonstrated that there exist real numbers  $\varepsilon_2 \in (0, 1)$ ,  $K_2 \in [1, \infty)$  (independent of  $\theta, x, x'$ ) such that the first two inequalities in (153) hold. In what follows in the proof of (i), we show that  $\partial_\theta^\alpha q_\theta(y|x)$  exists. We also demonstrate that there exists a real number  $K_3 \in [1, \infty)$  (independent of  $\theta, x, y$ ) such that the last two inequalities in (153) hold.

Relying on the same arguments as in the proof of Lemma 8.1, it can be shown that  $\partial_\theta^\alpha \tilde{C}_\theta(y|x)$ ,  $\partial_\theta^\alpha \tilde{C}_\theta'(y|x)$ ,  $\partial_\theta^\alpha V_\theta(y|x)$ ,  $\partial_\theta^\alpha v_\theta(y|x)$  exist and are continuous in  $(\theta, x, y)$  on  $\tilde{\Theta} \times \mathcal{X} \times \mathcal{Y}$ . Using the same arguments, it can be demonstrated that  $\partial_\theta^\alpha \tilde{D}_\theta(x)$ ,  $\partial_\theta^\alpha \tilde{D}_\theta'(x)$  exist and are continuous in  $(\theta, x)$  on  $\tilde{\Theta} \times \mathcal{X}$ . Since  $\Theta$  is bounded and  $\text{cl}\Theta \subset \tilde{\Theta}$ , Assumptions 4.1, 4.3, 4.5 imply that there exist real numbers  $\delta \in (0, 1)$ ,  $\tilde{C}_1 \in [1, \infty)$  (independent of  $\theta, x, \beta$ ) such that

$$\left| \tilde{D}_\theta(x) \right| \geq \delta, \quad \max \left\{ \left| \partial_\theta^\beta \tilde{D}_\theta(x) \right|, \left\| \partial_\theta^\beta \tilde{D}_\theta'(x) \right\| \right\} \leq \tilde{C}_1 \quad (155)$$

for  $\beta \in \mathbb{N}_0^d$ ,  $|\beta| \leq p$ . The same arguments also yield that there exists a real number  $\tilde{C}_2 \in [1, \infty)$  (independent of  $\theta, x, y, \gamma$ ) such that

$$\left\| \tilde{C}_\theta'(y|x) \right\| \leq \tilde{C}_2(1 + \|y\|), \quad \left\| \partial_\theta^\gamma \tilde{C}_\theta'(y|x) \right\| \leq \tilde{C}_2 \quad (156)$$

for  $\gamma \in \mathbb{N}_0^d \setminus \{0\}$ ,  $|\gamma| \leq p$ .

Let  $\tilde{C}_3 = 2^p \tilde{C}_1 \tilde{C}_2$ . Owing to Leibniz formula and (155), (156), we have

$$\begin{aligned} \left\| \partial_\theta^\alpha \tilde{C}_\theta(y|x) \right\| &\leq \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \left\| \partial_\theta^\beta \tilde{D}_\theta'(x) \right\| \left\| \partial_\theta^{\alpha-\beta} \tilde{C}_\theta'(y|x) \right\| \leq \tilde{C}_1 \tilde{C}_2 \left( 1 + \|y\| + \sum_{\substack{\beta \in \mathbb{N}_0^d \setminus \{0\} \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \right) \\ &\leq 2^{|\alpha|} \tilde{C}_1 \tilde{C}_2 (1 + \|y\|) \\ &\leq \tilde{C}_3 (1 + \|y\|). \end{aligned}$$

Consequently, Lemma A2.1 (see Appendix 2) and (155) imply that there exists a real number  $\tilde{C}_4 \in [1, \infty)$  (independent of  $\theta, x, y, \alpha$ ) such that

$$\left\| \partial_\theta^\alpha V_\theta(y|x) \right\| \leq \tilde{C}_4 (1 + \|y\|). \quad (157)$$

Then, Lemma A1.1 (see Appendix 1) and Assumption 4.5, yield that there exists a real number  $\tilde{C}_5 \in [1, \infty)$  (independent of  $\theta, x, y, \alpha$ ) such that

$$v_\theta(y|x) \leq \tilde{C}_5, \quad |\partial_\theta^\alpha v_\theta(y|x)| \leq \tilde{C}_5 v_\theta(y|x) (1 + \|y\|)^{2|\alpha|}. \quad (158)$$

Moreover, due to Assumptions 4.1, 4.2, the sign of  $\tilde{D}_\theta(x)$  is constant in  $\theta$  on each connected component of  $\Theta$ . Since  $\Theta$  is open, all connected components of  $\Theta$  are open, too. As  $\bar{v}_\theta(x) = |\tilde{D}_\theta(x)|$  (due to Assumption 4.5 and  $\mathcal{Y} = \mathbb{R}^{d_y}$ ), we conclude that  $\partial_\theta^\alpha \bar{v}_\theta(x)$  exists. Using (155), we also deduce

$$\bar{v}_\theta(x) \geq \delta, \quad |\partial_\theta^\alpha \bar{v}_\theta(x)| = \left| \partial_\theta^\alpha \tilde{D}_\theta(x) \right| \leq \tilde{C}_1. \quad (159)$$

Consequently, Lemma A2.1 implies that  $\partial_\theta^\alpha q_\theta(y|x)$  exists. The same lemma, Assumption 4.5 and (158), (159) also yield that there exists a real number  $K_3 \in [1, \infty)$  (independent of  $\theta, x, y, \alpha$ ) such that the last two inequalities in (153) hold.

(ii) Let  $\tilde{C}_6 = 5L_0\tilde{C}_1\tilde{C}_4^2$ ,  $K_4 = K_0\tilde{C}_6$ . Owing to Assumption 4.6 and (155), (157), we have

$$\begin{aligned} \log q_\theta(y|x) &= \log v_\theta(y|x) - \log \bar{v}_\theta(x) = \log s(V_\theta(y|x)) - \log |\tilde{D}_\theta(x)| \\ &\geq -L_0(1 + \|V_\theta(y|x)\|)^2 - \tilde{C}_1 \\ &\geq -4L_0\tilde{C}_4^2(1 + \|y\|)^2 - \tilde{C}_1 \\ &\geq -\tilde{C}_6(1 + \|y\|)^2. \end{aligned} \quad (160)$$

Moreover, due to Assumption 4.5, we have

$$\log q_\theta(y|x) \leq \log K_0 \leq K_0(1 + \|y\|)^2. \quad (161)$$

Combining (160), (161), we conclude that (154) holds.  $\square$

**Proof of Corollary 4.2.** Throughout the proof, the following notation is used.  $\varepsilon, \tilde{C}_1, \tilde{C}_2, \tilde{C}_3$  are the real numbers defined by  $\varepsilon = \min\{\varepsilon_2, K_2^{-1}\}$ ,  $\tilde{C}_1 = 2K_2K_3\varepsilon_2^{-2}$ ,  $\tilde{C}_2 = K_2K_3$ ,  $\tilde{C}_3 = K_3K_4(1 + |\log \mu(\mathcal{X})|)$  ( $\varepsilon_2, K_2, K_3, K_4$  are specified in Lemma 8.2).  $r, u, v$  are the real numbers specified in Assumptions 2.6, 3.2.  $\psi(y)$ ,  $\phi(x, y)$ ,  $\varphi(y)$  and  $\mu_\theta(dx|y)$  are the functions and the measure defined by

$$\psi(y) = \tilde{C}_1(1 + \|y\|)^2, \quad \phi(x, y) = \tilde{C}_2, \quad \varphi(y) = \tilde{C}_3(1 + \|y\|)^2, \quad \mu_\theta(B|y) = \int_B q_\theta(y|x)\mu(dx)$$

for  $\theta \in \Theta$ ,  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ ,  $B \in \mathcal{B}(\mathcal{X})$  ( $\mu(dx)$  is specified in Subsection 2.1).  $r_\theta(y, x'|x)$  has the same meaning as in (1), while  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  are defined in (20).  $\theta$  is any element of  $\Theta$ , while  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .  $x, x'$  are any elements of  $\mathcal{X}$ , while  $y$  is any element of  $\mathcal{Y}$ .

(i) Owing to Lemma 8.2, we have

$$\varepsilon_2 q_\theta(y|x') \leq r_\theta(y, x'|x) \leq K_2 q_\theta(y|x') \leq K_2 K_3. \quad (162)$$

Consequently, we get

$$\begin{aligned} \int_B r_\theta(y, x'|x)\mu(dx') &\geq \varepsilon_2 \int_B q_\theta(y|x')\mu(dx') \geq \varepsilon \mu_\theta(B|y), \\ \int_B r_\theta(y, x'|x)\mu(dx') &\leq K_2 \int_B q_\theta(y|x')\mu(dx') \leq \frac{1}{\varepsilon} \mu_\theta(B|y) \end{aligned}$$

for  $B \in \mathcal{B}(\mathcal{X})$ . We also get

$$r_\theta(y, x'|x) \leq \tilde{C}_2 = \phi(y, x'), \quad \int \phi(y, x)\mu(dx) = \tilde{C}_2\mu(\mathcal{X}) < \infty.$$

Hence, Assumptions 2.1, 2.3 hold for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20).

Due to Leibniz formula and Lemma 8.2, we have

$$\begin{aligned} |\partial_\theta^\alpha r_\theta(y, x'|x)| &\leq \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} \left| \partial_\theta^\beta q_\theta(y|x') \right| \left| \partial_\theta^{\alpha-\beta} p_\theta(x'|x) \right| \leq K_2 K_3 q_\theta(y|x') \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \binom{\alpha}{\beta} (1 + \|y\|)^{2|\beta|} \\ &\leq 2^{|\alpha|} K_2 K_3 q_\theta(y|x') (1 + \|y\|)^{2|\alpha|}. \end{aligned}$$

Then, (162) implies

$$|\partial_\theta^\alpha r_\theta(y, x'|x)| \leq 2^{|\alpha|} K_2 K_3 \varepsilon_2^{-1} (1 + \|y\|)^{2|\alpha|} r_\theta(y, x'|x) \leq (\psi(y))^{|\alpha|} r_\theta(y, x'|x).$$

Thus, Assumption 2.2 holds for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20). Consequently, all conclusions of Theorems 2.1, 2.2 are true for the model introduced in Section 4.

(ii) Owing to (22), we have

$$\int \varphi(x, y) \psi^r(y) Q(x, dy) \leq \tilde{C}_1^r \sup_{x' \in \mathcal{X}} \int \varphi(x', y) (1 + \|y\|)^{2r} Q(x', dy) < \infty.$$

Hence, in addition to Assumptions 2.1 – 2.3, Assumptions 2.4 – 2.6 also hold for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20). Therefore, all conclusions of Theorem 2.3 are true for the model introduced in Section 4.

(iii) Owing to Lemma 8.2, we have

$$\mu_\theta(\mathcal{X}|y) = \int q_\theta(y|x) \mu(dx) \leq K_3 \mu(\mathcal{X}). \quad (163)$$

Due to the same lemma and Jensen inequality, we also have

$$\log \mu_\theta(\mathcal{X}|y) \geq \log \mu(\mathcal{X}) + \frac{1}{\mu(\mathcal{X})} \int \log q_\theta(y|x) \mu(dx) \geq -|\log \mu(\mathcal{X})| - K_4 (1 + \|y\|)^2. \quad (164)$$

Combining (163), (164), we get

$$|\log \mu_\theta(\mathcal{X}|y)| \leq K_3 |\log \mu(\mathcal{X})| + K_4 (1 + \|y\|)^2 \leq \tilde{C}_3 (1 + \|y\|)^2 = \varphi(y).$$

Moreover, (23) implies

$$\int \psi^v(y) Q(x, dy) \leq \tilde{C}_1^v \sup_{x' \in \mathcal{X}} \int (1 + \|y\|)^{2v} Q(x', dy) < \infty.$$

As  $v \geq u + 1$ , (23) also yields

$$\int \varphi(y) \psi^u(y) Q(x, dy) \leq \tilde{C}_1^u \tilde{C}_3 \sup_{x' \in \mathcal{X}} \int (1 + \|y\|)^{2(u+1)} Q(x', dy) < \infty.$$

Thus, in addition to Assumptions 2.1 – 2.3, Assumptions 2.4, 3.1, 3.2 also hold for  $p_\theta(x'|x)$ ,  $q_\theta(y|x)$  specified in (20). Consequently, all conclusions of Theorem 3.1 are true for the model introduced in Section 4.  $\square$

## Appendix 1

In this section, we present auxiliary results crucially important for the proof of Corollaries 4.1 and 4.2. Let  $\Theta$  and  $d$  have the same meaning as in Subsection 2.1. Moreover, let  $\mathcal{Z}$  be an open set in  $\mathbb{R}^{d_z}$ , where  $d_z \geq 1$  is an integer. We consider here functions  $f_\theta$  and  $g(z)$  mapping  $\theta \in \Theta$ ,  $z \in \mathcal{Z}$  to  $\mathcal{Z}$  and  $\mathbb{R}$  (respectively). We also consider function  $h_\theta$  defined by  $h_\theta = g(f_\theta)$  for  $\theta \in \Theta$ . The analysis carried out in this section relies on the following assumptions.

**Assumption A1.1.**  $f_\theta$  and  $g(z)$  are  $p$ -times differentiable on  $\Theta$  and  $\mathcal{Z}$  (respectively), where  $p \geq 1$  is an integer.

**Assumption A1.2.** *There exist a real number  $K \in [1, \infty)$  and a function  $\phi_\theta$  mapping  $\theta \in \Theta$  to  $[1, \infty)$  such that*

$$\max \{ \|f_\theta\|, \|\partial_\theta^\alpha f_\theta\| \} \leq \phi_\theta, \quad |\partial^\beta g(z)| \leq K |g(z)| (1 + \|z\|)^{|\beta|}$$

for all  $\theta \in \Theta$ ,  $z \in \mathcal{Z}$  and any multi-indices  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ ,  $\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\}$  satisfying  $|\alpha| \leq p$ ,  $|\beta| \leq p$ .

Throughout this section, the following notation is used. For  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ,  $n_\alpha$  and  $m_\alpha$  are the integers defined by

$$n_\alpha = (\alpha_1 + 1) \cdots (\alpha_d + 1) - 1, \quad m_\alpha = \binom{n_\alpha}{|\alpha|}.$$

For  $\theta \in \Theta$ ,  $1 \leq k \leq d_z$ ,  $f_{\theta,k}$  is the  $k$ -th component of  $f_\theta$ . For the same  $\theta$  and  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ ,  $|\alpha| \leq p$ ,  $F_{\theta,\alpha}$  is the  $n_\alpha$ -dimensional vector whose components are derivatives  $\{\partial_\theta^\beta f_{\theta,k} : \beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\}, \beta \leq \alpha, 1 \leq k \leq d_z\}$ . In  $F_{\theta,\alpha}$ , the components are ordered lexicographically in  $(k, \beta)$ .

**Lemma A1.1.** (i) *Let Assumption A1.1 hold. Then,  $h_\theta$  is  $p$ -times differentiable on  $\Theta$ . Moreover, the first and higher-order derivatives of  $h_\theta$  admit representation*

$$\partial_\theta^\alpha h_\theta = \sum_{\substack{\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\} \\ |\beta| \leq |\alpha|}} \partial^\beta g(f_\theta) P_{\alpha,\beta}(F_{\theta,\alpha}) \quad (165)$$

for all  $\theta \in \Theta$  and any multi-index  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$  satisfying  $|\alpha| \leq p$ . Here,  $P_{\alpha,\beta} : \mathbb{R}^{n_\alpha} \rightarrow \mathbb{R}$  is a polynomial of degree up to  $|\alpha|$  whose coefficients are independent of  $\theta$  and depend only on  $\alpha, \beta$ .

(ii) *Let Assumptions A1.1 and A1.2 hold. Then, there exists a real number  $L \in [1, \infty)$  such that*

$$|\partial_\theta^\alpha h_\theta| \leq L |h_\theta| \phi_\theta^{2|\alpha|}$$

for all  $\theta \in \Theta$  and any multi-index  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$  satisfying  $|\alpha| \leq p$ .

*Proof.* (i) This part of lemma is proved by induction in  $|\alpha|$ . It is straightforward to show that  $\partial_\theta^\alpha h_\theta$  exists and satisfies (165) for all  $\theta \in \Theta$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = 1$ . Now, the induction hypothesis is formulated. Let  $1 \leq l < p$  be an integer. Suppose that  $\partial_\theta^\alpha h_\theta$  exists and satisfies (165) for each  $\theta \in \Theta$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq l$ . Then, to show (i), it is sufficient to demonstrate that  $\partial_\theta^\alpha h_\theta$  exists and satisfies (165) for all  $\theta \in \Theta$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = l + 1$ .

Let  $\theta$  be any element of  $\Theta$ , while  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| = l + 1$ . Then, there exists  $e \in \mathbb{N}_0^d$  such that  $e \leq \alpha$ ,  $|e| = 1$ . As  $|\alpha - e| = |\alpha| - 1 = l$ , the induction hypothesis yields

$$\partial_\theta^{\alpha-e} h_\theta = \sum_{\substack{\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\} \\ |\beta| \leq l}} \partial^\beta g(f_\theta) P_{\alpha-e,\beta}(F_{\theta,\alpha-e}). \quad (166)$$

Since  $l < p$ , the right-hand side of (166) involves only the derivatives of  $f_\theta, g(z)$  of the order up to  $p - 1$ . Then, Assumption A1.1 implies that  $\partial_\theta^\alpha h_\theta = \partial_\theta^e (\partial_\theta^{\alpha-e} h_\theta)$  exist and satisfies

$$\begin{aligned} \partial_\theta^\alpha h_\theta &= \sum_{\substack{\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\} \\ |\beta| \leq l}} \partial^\beta g(f_\theta) \partial_\theta^e P_{\alpha-e,\beta}(F_{\theta,\alpha-e}) + \sum_{k=1}^{d_z} \sum_{\substack{\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\} \\ |\beta| \leq l}} \partial^{\beta+e_k} g(f_\theta) \partial_\theta^e f_{\theta,k} P_{\alpha-e,\beta}(F_{\theta,\alpha-e}) \\ &= \sum_{\substack{\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\} \\ |\beta| \leq l}} \partial^\beta g(f_\theta) \partial_\theta^e P_{\alpha-e,\beta}(F_{\theta,\alpha-e}) + \sum_{k=1}^{d_z} \sum_{\substack{\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\} \\ e_k \leq \beta, |\beta| \leq l+1}} \partial^\beta g(f_\theta) \partial_\theta^e f_{\theta,k} P_{\alpha-e,\beta-e_k}(F_{\theta,\alpha-e}), \end{aligned} \quad (167)$$

where  $e_k$  is the  $k$ -th standard unit vector in  $\mathbb{N}_0^{d_z}$ . Moreover, terms

$$\partial_\theta^e P_{\alpha-e,\beta}(F_{\theta,\alpha-e}), \quad \partial_\theta^e f_{\theta,k} P_{\alpha-e,\beta-e_k}(F_{\theta,\alpha-e})$$

are polynomial in derivatives  $\{\partial_\theta^\gamma f_{\theta,j} : \gamma \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}, \gamma \leq \alpha, 1 \leq j \leq d_z\}$ . Apparently, the order of these polynomials is up to  $|\alpha - e| + 1 = |\alpha|$ , while the corresponding coefficients are independent of  $\theta$  and depend only on  $\alpha, \beta$ . Therefore, the right-hand side of (167) admits representation (165). Hence, the same holds for  $\partial_\theta^\alpha h_\theta$ .

(ii) Let  $\tilde{C}_{\alpha,\beta}$  be the maximum absolute value of the coefficients of polynomial  $P_{\alpha,\beta}(\cdot)$ , where  $\alpha \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}$ ,  $\beta \in \mathbb{N}_0^{d_z} \setminus \{\mathbf{0}\}$ ,  $|\alpha| \leq p$ ,  $|\beta| \leq |\alpha|$ . As the number of different power terms in  $P_{\alpha,\beta}(\cdot)$  is at most  $m_\alpha$ , Assumption A1.2 and (i) yield

$$\begin{aligned} |\partial^\beta g(f_\theta) P_{\alpha,\beta}(F_{\theta,\alpha})| &\leq K \tilde{C}_{\alpha,\beta} m_\alpha |g(f_\theta)| (1 + \|f_\theta\|)^{|\beta|} \phi_\theta^{|\alpha|} \leq K \tilde{C}_{\alpha,\beta} m_\alpha |h_\theta| (1 + \phi_\theta)^{|\alpha|} \phi_\theta^{|\alpha|} \\ &\leq 2^{|\alpha|} K \tilde{C}_{\alpha,\beta} m_\alpha |h_\theta| \phi_\theta^{2|\alpha|}. \end{aligned}$$

Then, using (i) again, we conclude that there exists a real number  $L \in [1, \infty)$  with the properties specified in the lemma's statement.  $\square$

## Appendix 2

As the previous section, this section provides auxiliary results relevant for the proof of Corollaries 4.1 and 4.2. Let  $\Theta$  and  $d$  have the same meaning as in Subsection 2.1. We consider here functions  $f_\theta$  and  $g_\theta$  mapping  $\theta \in \Theta$  to  $\mathbb{R}$  and  $\mathbb{R} \setminus \{0\}$  (respectively). We also consider function  $h_\theta$  defined by  $h_\theta = f_\theta/g_\theta$  for  $\theta \in \Theta$ . The results presented in this section rely on the following assumptions.

**Assumption A2.1.**  $f_\theta$  and  $g_\theta$  are  $p$ -times differentiable on  $\Theta$ , where  $p \geq 1$  is an integer.

**Assumption A2.2.** There exist functions  $\phi_\theta$  and  $\psi_\theta$  mapping  $\theta \in \Theta$  to  $[1, \infty)$  such that

$$|\partial_\theta^\alpha f_\theta| \leq |f_\theta| \phi_\theta^{|\alpha|}, \quad |\partial_\theta^\alpha g_\theta| \leq \psi_\theta$$

for all  $\theta \in \Theta$  and any multi-index  $\alpha \in \mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .

Throughout this section, we use the following notation. For  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ,  $n_\alpha$  and  $m_\alpha$  are the integers defined by

$$n_\alpha = (\alpha_1 + 1) \cdots (\alpha_d + 1), \quad m_\alpha = \binom{n_\alpha}{|\alpha|}.$$

For  $\theta \in \Theta$  and  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ ,  $G_{\theta,\alpha}$  is the  $n_\alpha$ -dimensional vector whose components are derivatives  $\{\partial_\theta^\beta g_\theta : \beta \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}, \beta \leq \alpha, 1 \leq k \leq d_z\}$ . In  $G_{\theta,\alpha}$ , the components are ordered lexicographically in  $\beta$ .

**Lemma A2.1.** (i) Let Assumption A2.1 hold. Then,  $h_\theta$  is  $p$ -times differentiable on  $\Theta$ . Moreover, the first and higher-order derivatives of  $h_\theta$  admit representation

$$\partial_\theta^\alpha h_\theta = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \frac{\partial_\theta^\beta f_\theta P_{\alpha,\beta}(G_{\theta,\alpha})}{g_\theta^{|\alpha|+1}} \quad (168)$$

for all  $\theta \in \Theta$  and any multi-index  $\alpha \in \mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ . Here,  $P_{\alpha,\beta} : \mathbb{R}^{n_\alpha} \rightarrow \mathbb{R}$  is a polynomial of the degree up to  $|\alpha|$  whose coefficients are independent of  $\theta$  and depend only on  $\alpha, \beta$ .

(ii) Let Assumptions A2.1 and A2.2 hold. Then, there exists a real number  $K \in [1, \infty)$  such that

$$|\partial_\theta^\alpha h_\theta| \leq K \left| \frac{f_\theta}{g_\theta} \right| \left( \frac{\phi_\theta \psi_\theta}{|g_\theta|} \right)^{|\alpha|}$$

for all  $\theta \in \Theta$  and any multi-index  $\alpha \in \mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ .

*Proof.* (i) This part of lemma is proved by induction in  $|\alpha|$ . It is straightforward to show that  $\partial_\theta^\alpha h_\theta$  exists and satisfies (168) for all  $\theta \in \Theta$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \in \{0, 1\}$ . Now, the induction hypothesis is formulated. Let  $1 \leq l < p$  be an integer. Suppose that  $\partial_\theta^\alpha h_\theta$  exists and satisfies (168) for each  $\theta \in \Theta$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq l$ . Then, to show (i), it is sufficient to demonstrate that  $\partial_\theta^\alpha h_\theta$  exists and satisfies (168) for all  $\theta \in \Theta$ ,  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| = l + 1$ .

Let  $\theta$  be any element of  $\Theta$ , while  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| = l + 1$ . Then, there exists  $e \in \mathbb{N}_0^d$  such that  $e \leq \alpha$ ,  $|e| = 1$ . As  $|\alpha - e| = |\alpha| - 1 = l$ , the induction hypothesis yields

$$\partial_\theta^{\alpha-e} h_\theta = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha-e}} \frac{\partial_\theta^\beta f_\theta P_{\alpha-e, \beta}(G_\theta, \alpha-e)}{g_\theta^{|\alpha|}}. \quad (169)$$

Since  $l < p$ , the right-hand side of (169) involves only the derivatives of  $f_\theta$ ,  $g_\theta$  of the order up to  $p - 1$ . Then, Assumption A2.1 implies that  $\partial_\theta^\alpha h_\theta = \partial_\theta^e (\partial_\theta^{\alpha-e} h_\theta)$  exist and satisfies

$$\begin{aligned} \partial_\theta^\alpha h_\theta &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha-e}} \frac{\partial_\theta^{\beta+e} f_\theta P_{\alpha-e, \beta}(G_\theta, \alpha-e) + \partial_\theta^\beta f_\theta \partial_\theta^e P_{\alpha-e, \beta}(G_\theta, \alpha-e)}{g_\theta^{|\alpha|}} - |\alpha| \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha-e}} \frac{\partial_\theta^\beta f_\theta \partial_\theta^e g_\theta P_{\alpha-e, \beta}(G_\theta, \alpha-e)}{g_\theta^{|\alpha|+1}} \\ &= \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha-e}} \frac{\partial_\theta^\beta f_\theta (g_\theta \partial_\theta^e P_{\alpha-e, \beta}(G_\theta, \alpha-e) - |\alpha| \partial_\theta^\beta g_\theta P_{\alpha-e, \beta}(G_\theta, \alpha-e))}{g_\theta^{|\alpha|+1}} + \sum_{\substack{\beta \in \mathbb{N}_0^d \\ e \leq \beta \leq \alpha}} \frac{\partial_\theta^\beta f_\theta P_{\alpha-e, \beta-e}(G_\theta, \alpha-e)}{g_\theta^{|\alpha|}} \end{aligned} \quad (170)$$

Moreover, terms

$$\partial_\theta^e g_\theta P_{\alpha-e, \beta}(G_\theta, \alpha-e), \quad g_\theta \partial_\theta^e P_{\alpha-e, \beta}(G_\theta, \alpha-e), \quad g_\theta P_{\alpha-e, \beta-e}(G_\theta, \alpha-e)$$

are polynomial in derivatives  $\{\partial_\theta^\gamma g_\theta : \gamma \in \mathbb{N}_0^d, \gamma \leq \alpha\}$ . Apparently, the order of these polynomials is up to  $|\alpha - e| + 1 = |\alpha|$ , while the corresponding coefficients are independent of  $\theta$  and depend only on  $\alpha, \beta$ . Therefore, the right-hand side of (170) admits representation (168). Hence, the same holds for  $\partial_\theta^\alpha h_\theta$ .

(ii) Let  $\tilde{C}_{\alpha, \beta}$  be the maximum absolute value of the coefficients of polynomial  $P_{\alpha, \beta}(\cdot)$ , where  $\alpha, \beta \in \mathbb{N}_0^d$ ,  $\alpha \leq \beta$ . As the number of different power terms in  $P_{\alpha, \beta}(\cdot)$  is at most  $m_\alpha$ , Assumption A2.2 and (i) yield

$$\left| \frac{\partial_\theta^\beta f_\theta P_{\alpha, \beta}(G_\theta, \alpha)}{g_\theta^{|\alpha|+1}} \right| \leq \tilde{C}_{\alpha, \beta} m_\alpha \phi_\theta^{|\beta|} \left| \frac{f_\theta}{g_\theta} \right| \left( \frac{|\psi_\theta|}{|g_\theta|} \right)^{|\alpha|} \leq \tilde{C}_{\alpha, \beta} m_\alpha \left| \frac{f_\theta}{g_\theta} \right| \left( \frac{\phi_\theta \psi_\theta}{|g_\theta|} \right)^{|\alpha|}$$

for all  $\theta \in \Theta$  and any  $\alpha, \beta \in \mathbb{N}_0^d \setminus \{0\}$ ,  $\beta \leq \alpha$ . Then, using (i) again, we conclude that there exists a real number  $K \in [1, \infty)$  with the properties specified in the lemma's statement.  $\square$

## Appendix 3

In this section, we present auxiliary results which Proposition 7.1 and Theorem 2.1 crucially rely on. Let  $\Theta$  and  $d$  have the same meaning as in Section 2. Moreover, let  $\mathcal{Z}$  be a Borel set in  $\mathbb{R}^{d_z}$ , where  $d_z \geq 1$  is an integer. We consider here functions  $F_\theta(z)$  and  $g_\theta$  mapping  $\theta \in \Theta$ ,  $z \in \mathcal{Z}$  to  $\mathbb{R}$  and  $\mathbb{R} \setminus \{0\}$  (respectively). We also consider non-negative measure  $\mu(dz)$  on  $\mathcal{Z}$ . The analysis carried out in this section relies on the following assumptions.

**Assumption A3.1.**  $F_\theta(z)$  and  $g_\theta$  are  $p$ -times differentiable in  $\theta$  for each  $\theta \in \Theta$ ,  $z \in \mathcal{Z}$ , where  $p \geq 1$ .

**Assumption A3.2.** There exists a function  $\phi : \mathcal{Z} \rightarrow [1, \infty)$  such that

$$|\partial_\theta^\alpha F_\theta(z)| \leq \phi(z), \quad \int \phi(z') \mu(dz') < \infty$$

for all  $\theta \in \Theta$ ,  $z \in \mathcal{Z}$  and any multi-index  $\alpha \in \mathbb{N}_0^d$ ,  $|\alpha| \leq p$ .



Throughout this section, we use the following notation.  $f_\theta, h_\theta, H_\theta(z)$  are the functions defined by

$$f_\theta = \int F_\theta(z')\mu(dz'), \quad h_\theta = \frac{f_\theta}{g_\theta}, \quad H_\theta(z) = \frac{F_\theta(z)}{g_\theta}$$

for  $\theta \in \Theta, z \in \mathcal{Z}$ .  $\xi_\theta(dz)$  and  $\zeta_\theta(dz)$  are the signed measures on  $\mathcal{Z}$  defined by

$$\xi_\theta(B) = \int_B F_\theta(z)\mu(dz), \quad \zeta_\theta(B) = \int_B H_\theta(z)\mu(dz)$$

for  $B \in \mathcal{B}(\mathcal{Z})$ .  $\xi_\theta^\alpha(dz)$  and  $\zeta_\theta^\alpha(dz)$  are the signed measures on  $\mathcal{Z}$  defined by

$$\xi_\theta^\alpha(B) = \int_B \partial_\theta^\alpha F_\theta(z)\mu(dz), \quad \zeta_\theta^\alpha(B) = \int_B \partial_\theta^\alpha H_\theta(z)\mu(dz)$$

for  $\alpha \in \mathbb{N}_0^d, |\alpha| \leq p$ .

**Lemma A3.1.** *Let Assumptions A3.1 and A3.2 hold. Then, the following is true.*

(i)  $f_\theta$  and  $g_\theta$  are well-defined for each  $\theta \in \Theta$ . Moreover,  $f_\theta$  and  $g_\theta$  are  $p$ -times differentiable and satisfy

$$\partial_\theta^\alpha f_\theta = \int \partial_\theta^\alpha F_\theta(z)\mu(dz), \quad \partial_\theta^\alpha h_\theta = \int \partial_\theta^\alpha H_\theta(z)\mu(dz) \quad (171)$$

for all  $\theta \in \Theta$  and any multi-index  $\alpha \in \mathbb{N}_0^d, |\alpha| \leq p$ .

(ii)  $\xi_\theta(B), \zeta_\theta(B), \xi_\theta^\alpha(B)$  and  $\zeta_\theta^\alpha(B)$  are well-defined for each  $\theta \in \Theta, B \in \mathcal{B}(\mathcal{Z})$ . Moreover,  $\xi_\theta(B)$  and  $\zeta_\theta(B)$  are  $p$ -times differentiable (in  $\theta$ ) and satisfy

$$\partial_\theta^\alpha \xi_\theta(B) = \xi_\theta^\alpha(B), \quad \partial_\theta^\alpha \zeta_\theta(B) = \zeta_\theta^\alpha(B) \quad (172)$$

for all  $\theta \in \Theta, B \in \mathcal{B}(\mathcal{Z})$  and any multi-index  $\alpha \in \mathbb{N}_0^d, |\alpha| \leq p$ .

*Proof.* Let  $\theta$  be any element of  $\Theta$ , while  $\alpha$  is any multi-index in  $\mathbb{N}_0^d$  satisfying  $|\alpha| \leq p$ . Moreover, let  $z$  be any element of  $\mathcal{Z}$ , while  $B$  is any element of  $\mathcal{B}(\mathcal{Z})$ . Owing to Assumptions A3.1, A3.2,  $f_\theta, \xi_\theta(B), \xi_\theta^\alpha(B)$  are well-defined. Consequently,  $h_\theta, \zeta_\theta(B)$  are also well-defined. Moreover, due to the dominated convergence theorem and Assumptions A3.1, A3.2,  $f_\theta, \xi_\theta(B)$  are  $p$ -times differentiable in  $\theta$  on  $\Theta$  and satisfy the first part of (171), (172). Therefore,  $h_\theta, \zeta_\theta(B)$  are also  $p$ -times differentiable in  $\theta$  on  $\Theta$ .

Using Lemma A2.1, we conclude that  $\partial_\theta^\alpha h_\theta, \partial_\theta^\alpha H_\theta(z), \partial_\theta^\alpha \zeta_\theta(B)$  admit the following representation:

$$\partial_\theta^\alpha h_\theta = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \frac{G_\theta^{\alpha, \beta} \partial_\theta^\beta f_\theta}{(g_\theta)^{|\alpha|+1}}, \quad \partial_\theta^\alpha H_\theta(x) = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \frac{G_\theta^{\alpha, \beta} \partial_\theta^\beta F_\theta(z)}{(g_\theta)^{|\alpha|+1}}, \quad \partial_\theta^\alpha \zeta_\theta(B) = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \frac{G_\theta^{\alpha, \beta} \partial_\theta^\beta \xi_\theta(B)}{(g_\theta)^{|\alpha|+1}}, \quad (173)$$

where  $G_\theta^{\alpha, \beta}$  is a polynomial function of derivatives  $\{\partial_\theta^\gamma g_\theta : \gamma \in \mathbb{N}_0, \gamma \leq \alpha\}$ . Owing to (173) and the first part of (171), we have

$$\partial_\theta^\alpha h_\theta = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \frac{G_\theta^{\alpha, \beta}}{(g_\theta)^{|\alpha|+1}} \int \partial_\theta^\beta F_\theta(z)\mu(dz) = \int \partial_\theta^\alpha H_\theta(z)\mu(dz).$$

Similarly, due to (173) and the first part of (172), we have

$$\partial_\theta^\alpha \zeta_\theta(B) = \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq \alpha}} \frac{G_\theta^{\alpha, \beta}}{(g_\theta)^{|\alpha|+1}} \int_B \partial_\theta^\beta F_\theta(z)\mu(dz) = \int_B \partial_\theta^\alpha H_\theta(z)\mu(dz) = \zeta_\theta^\alpha(B).$$

Hence,  $\partial_\theta^\alpha h_\theta, \zeta_\theta^\alpha(B)$  are well-defined and satisfy the second part of (171), (172).  $\square$

## References

- [1] R. Atar and O. Zeitouni, *Exponential stability of nonlinear filtering*, Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, 33 (1997), pp. 697 – 725.
- [2] P. J. Bickel, Y. Ritov, and T. Ryden, *Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models*, Annals of Statistics, 26 (1998), pp. 1614 – 1635.
- [3] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer-Verlag, 2005.
- [4] D. Crisan and B. Rozovskii, *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, 2011.
- [5] P. Del Moral and A. Guionnet, *On the stability of interacting processes with applications to filtering and genetic algorithms*, Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, 37 (2001), pp. 155 – 194.
- [6] P. Del Moral, A. Doucet, and S. S. Singh, *Uniform stability of a particle approximation of the optimal filter derivative*, SIAM Journal of Control and Optimization, 53 (2015), pp. 1278 – 1304.
- [7] R. Douc and C. Matias, *Asymptotics of the maximum likelihood estimator for general hidden Markov models*, Bernoulli, 7 (2001), pp. 381 – 420.
- [8] 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.
- [9] R. Douc, E. Moulines, and D. S. Stoffer, *Nonlinear Time Series: Theory, Methods and Applications with R Examples*, CRC Press, 2014.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] P. McCullagh, *Tensor Methods in Statistics*, CRC Press, 1987.
- [14] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd Edition, Cambridge University Press, 2009.
- [15] 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.
- [16] T. Ryden, *On recursive estimation for hidden Markov models*, Stochastic Processes and Their Applications 66 (1997), pp. 79 – 96.
- [17] V. B. Tadić, *Analyticity, convergence, and convergence rate of recursive maximum-likelihood estimation in hidden Markov models*, IEEE Transactions on Information Theory, 56 (2010), pp. 6406 – 6432.
- [18] 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.
- [19] V. B. Tadić and A. Doucet, *Asymptotic bias of stochastic gradient search*, Annals of Applied Probability, 27 (2017), pp. 3255–3304.

- [20] 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.
- [21] M. Taniguchi and Y. Kakizawa, *Asymptotic Theory of Statistical Inference for Time Series*, Springer, 2000.