Analyticity of Entropy Rates of Continuous-State Hidden Markov Models

Vladislav Z. B. Tadić and Arnaud Doucet

***Abstract*—The analyticity of the entropy and relative entropy rates of continuous-state hidden Markov models is studied here. Using the analytic continuation principle and the stability properties of the optimal filter, the analyticity of these rates is established for analytically parameterized models. The obtained results hold under relatively mild conditions and cover several useful classes of hidden Markov models. These results are relevant for several theoretically and practically important problems arising in statistical inference, system identification and information theory.**

***Index Terms*—Hidden Markov Models, Entropy Rate, Relative Entropy Rate, Log-Likelihood, Optimal Filter, Analytic Continuation.**

I. Introduction

Hidden Markov models are a powerful and versatile tool for statistical modeling of complex time-series data and stochastic dynamic systems. They can be described as a discrete-time Markov chain observed through imperfect, noisy observations of its states. Proposed in the seminal paper [1] over five decades ago, hidden Markov models have found many applications in very diverse areas such as acoustics and signal processing, image analysis and computer vision, automatic control, economics and finance, computational biology, genetics and bioinformatics. Owing to their theoretical and practical importance, various aspects of hidden Markov models have been thoroughly studied in a number of papers and books; see, e.g., [2], [6], [8] and references therein.

The entropy and relative entropy rates of hidden Markov models can be considered as an information-theoretic characterization of the asymptotic properties of these models. The entropy rate of a hidden Markov model can be interpreted as a measure of the average information revealed by the model through noisy observations of the states. The relative entropy rate between two hidden Markov models can be viewed as a measure of discrepancy between these models. The entropy rates of hidden Markov models and their analytical properties have recently gained significant attention in the information theory community. These properties and their links with statistical inference, system identification, stochastic optimization and information theory have been studied extensively in several papers [10] – [13], [14], [19], [20], [22], [23]. However, to the best of our knowledge, the existing results on the analytical properties of the entropy rates of hidden Markov models apply exclusively to scenarios where the hidden Markov chain takes values in a finite state-space. We establish here analytical properties of the entropy rates of continuous-state hidden Markov models.

For example, such results are useful when analyzing algorithms for parameter inference in hidden Markov models. In online settings, the unknown parameter is typically assessed using the recursive maximum likelihood method [23], [21]. In [23], it has been shown that the convergence and convergence rate of recursive maximum likelihood estimation in finite-state hidden Markov models is closely linked to the analyticity of the underlying average log-likelihood, i.e. of the underlying relative entropy rate. In view of recent results on stochastic gradient search [25], a similar link is expected to hold for continuous-state hidden Markov models. However, to apply the results of [25] to recursive maximum likelihood estimation in continuous-state hidden Markov models, it is necessary to establish the analyticity of the average log-likelihood for these models. Hence, one of the first and most important steps to carry out the asymptotic analysis of recursive maximum likelihood estimation in continuous-state hidden Markov models is to show the analyticity of the entropy rates of such models. The results presented here should provide a theoretical basis for this step.

In this paper, we study analytically parameterized continuous-state hidden Markov models (i.e., the models whose state transition kernel and the observation conditional distribution are analytic in the model parameters). Using mixing conditions on the model dynamics, we construct a geometrically ergodic analytic continuation of the state transition kernel and an exponentially stable analytic continuation of the optimal filter. Relying on these continuations and their asymptotic properties, we demonstrate that the entropy and relative entropy rates are analytic in the model parameters. The obtained results hold under relatively mild conditions and cover a broad and common class of state-space and continuous-state hidden Markov models. Moreover, these results generalize the existing results on the analyticity of entropy rates of finite-state hidden Markov models. Additionally, the results presented here are relevant for several important problems related to statistical inference, system identification and information theory. In [26], we use them to analyze the asymptotic properties of recursive maximum likelihood estimation in non-linear state-space models.

The rest of this paper is organized as follows. In Section II, the entropy rates of hidden Markov models are specified. In he same section, the main results are presented. Examples illustrating the main results are provided in Sections III and IV. In Sections V – VII, the main results are proved.

Vladislav Z. B. Tadić is with School of Mathematics, University of Bristol, Bristol, United Kingdom (e-mail: v.b.tadic@bristol.ac.uk).

Arnaud Doucet is with Department of Statistics, University of Oxford, Oxford, United Kingdom (e-mail: doucet@stats.ox.ac.uk).

II. Main Results

To define hidden Markov models and their entropy rates, we use the following notations. $(\Omega, \mathcal{F}, P)$ is a probability space. $d_x \geq 1$ and $d_y \geq 1$ are integers, while $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$ are Borel sets. $P(x, dx')$ is a transition kernel on $\mathcal{X}$, while $Q(x, dy)$ is a conditional probability measure on $\mathcal{Y}$ given $x \in \mathcal{X}$. Then, a hidden Markov model can be defined as the $\mathcal{X} \times \mathcal{Y}$-valued stochastic process $\{(X_n, Y_n)\}_{n\geq 0}$ which is defined on $(\Omega, \mathcal{F}, P)$ and satisfies

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

almost surely for $n \geq 0$ and any Borel set $B \subseteq \mathcal{X} \times \mathcal{Y}$. $\{X_n\}_{n\geq 0}$ are the unobservable states, while $\{Y_n\}_{n\geq 0}$ are the observations. Y_n can be interpreted as a noisy measurement of state X_n. States $\{X_n\}_{n\geq 0}$ form a Markov chain, while $P(x, dx')$ is their transition kernel. Conditionally on $\{X_n\}_{n\geq 0}$, state-observations $\{Y_n\}_{n\geq 0}$ are mutually independent, while $Q(X_n, dy)$ is the conditional distribution of Y_n given $X_{0:n}$. For more details on hidden Markov models, see [2], [6] and references therein.

Besides the model $\{(X_n, Y_n)\}_{n\geq 0}$, we also consider a parameterized family of hidden Markov models. To define such a family, we rely on the following notations. Let $d \geq 1$ be an integer, while $\Theta \subset \mathbb{R}^d$ is an open set. $\mathcal{P}(\mathcal{X})$ is the set of probability measures on $\mathcal{X}$. $\mu(dx)$ and $\nu(dy)$ are measures on $\mathcal{X}$ and $\mathcal{Y}$ (respectively) while $p_\theta(x'|x)$ 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

$$\int_{\mathcal{X}} p_\theta(x'|x)\mu(dx') = \int_{\mathcal{Y}} q_\theta(y|x)\nu(dy) = 1$$

for all $\theta \in \Theta$, $x \in \mathcal{X}$. A family of hidden Markov models can then be defined as a collection of $\mathcal{X} \times \mathcal{Y}$-valued stochastic processes $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ on $(\Omega, \mathcal{F}, P)$, parameterized by $\theta \in \Theta$, $\lambda \in \mathcal{P}(\mathcal{X})$ and satisfying

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

almost surely for $n \geq 0$ and any Borel set $B \subseteq \mathcal{X} \times \mathcal{Y}$.$\{X_n^{\theta,\lambda}\}_{n\geq 0}$ are the hidden states of this model, while $\{Y_n^{\theta,\lambda}\}_{n\geq 0}$ are the corresponding observations. $p_\theta(x'|x)$ is the transition density of the Markov chain $\{X_n^{\theta,\lambda}\}_{n\geq 0}$, while $q_\theta(y|X_n^{\theta,\lambda})$ is the conditional density of $Y_n^{\theta,\lambda}$ given $X_{0:n}^{\theta,\lambda}$. In the context of the identification of stochastic dynamical systems and parameter estimation in time-series models, $\{(X_n, Y_n)\}_{n\geq 0}$ is interpreted as the true system (or true model), while $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ is viewed as a candidate model for $\{(X_n, Y_n)\}_{n\geq 0}$.

To define the entropy rates, we introduce the transition density $r_\theta(y, x'|x)$ of $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$, i.e.,

$$r_\theta(y, x'|x) = q_\theta(y|x')p_\theta(x'|x),$$

and the density $q_\theta^n(y_{1:n}|\lambda)$ of $Y_{1:n}^{\theta,\lambda}$, i.e.,

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

where $y_{1:n} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ for $n \geq 1$. The (average) entropy of $Y_{1:n}^{\theta,\lambda}$ is given by

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

and the expected (average) log-likelihood $l_n(\theta, \lambda)$ given the model $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda}\right)$ is given by

$$l_n(\theta, \lambda) = E\left(\frac{1}{n}\log q_\theta^n\left(Y_{1:n}|\lambda\right)\right). \quad (2)$$

The entropy rate of model $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ (i.e., the entropy rate of stochastic process $\left\{Y_n^{\theta,\lambda}\right\}_{n\geq 0}$) can then be defined as the limit

$$\lim_{n\to\infty} h_n(\theta, \lambda).$$

Similarly, the relative entropy rate between models $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ and $\{(X_n, Y_n)\}_{n\geq 0}$ (i.e., the relative entropy rate between stochastic processes $\left\{Y_n^{\theta,\lambda}\right\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$) can be defined as the limit

$$-\lim_{n\to\infty}(l_n(\theta, \lambda) + h),$$

where h is the entropy rate of $\{Y_n\}_{n\geq 0}$ (provided h exists). In this context, the limit

$$\lim_{n\to\infty} l_n(\theta, \lambda)$$

can be viewed/referred to as the log-likelihood rate of $\{Y_n\}_{n\geq 0}$ given the model $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$. Entropy rate $\lim_{n\to\infty} h_n(\theta, \lambda)$ can be considered as a measure of the information revealed by the model $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ through its state-observations $\left\{Y_n^{\theta,\lambda}\right\}_{n\geq 0}$. Relative entropy rate $-\lim_{n\to\infty}(l_n(\theta, \lambda) + h)$ can be interpreted as a measure of discrepancy between the models $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ and $\{(X_n, Y_n)\}_{n\geq 0}$. The entropy rates of hidden Markov models are closely related to a number of important problems arising in engineering and statistics such as system identification, parameter estimation, model reduction and data compression. For example, in the recursive maximum likelihood approach to the identification of stochastic dynamical systems and parameter estimation in time-series models, the candidate model $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ providing the best approximation to the true model $\{(X_n, Y_n)\}_{n\geq 0}$ is selected through the minimization of $-\lim_{n\to\infty}(l_n(\theta, \lambda) + h)$ (i.e., through the maximization of $\lim_{n\to\infty} l_n(\theta, \lambda)$). For more details on the entropy rates and their applications, see [7], [9] and references therein.

We study here the rates $\lim_{n\to\infty} h_n(\theta, \lambda)$, $\lim_{n\to\infty} l_n(\theta, \lambda)$ and their analytical properties. To formulate the assumptions under which these rates are analyzed, we rely on the following notations. For $\eta \in \mathbb{C}^d$, $\|\eta\|$ denotes the Euclidean norm of η. For $\gamma \in (0,1)$, $V_\gamma(\Theta)$ is the open γ-vicinity of Θ in $\mathbb{C}^d$, i.e.,

$$V_\gamma(\Theta) = \{\eta \in \mathbb{C}^d : \exists \theta \in \Theta, \|\eta - \theta\| < \gamma\}.$$

Our analysis is based on the following assumptions.

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 finite measure $\lambda_\theta(dx|y)$ on $\mathcal{X}$ such that*

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

for all $x \in \mathcal{X}$ and any Borel set $B \subseteq \mathcal{X}$.

Assumption 2.2. *$r_\theta(y,x'|x)$ is real-analytic in θ for each $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Moreover, $r_\theta(y,x'|x)$ has a complex-valued continuation $\hat{r}_\eta(y,x'|x)$ with the following properties:*

(i) $\hat{r}_\eta(y,x'|x)$ maps $\eta \in \mathbb{C}^d$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ to $\mathbb{C}$.

(ii) $\hat{r}_\theta(y,x'|x) = r_\theta(y,x'|x)$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

(iii) There exists a real number $\delta \in (0,1)$ such that $\hat{r}_\eta(y,x'|x)$ is analytic in η for each $\eta \in V_\delta(\Theta)$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

(iv) There exists a function $\varphi_\eta(y)$ which maps $\eta \in \mathbb{C}^d$, $y \in \mathcal{Y}$ to $\mathbb{C}$, is analytic in η for each $\eta \in V_\delta(\Theta)$, $y \in \mathcal{Y}$ and satisfies

$$\varphi_\eta(y) \neq 0, \qquad |\hat{r}_\eta(y,x'|x)| \leq |\varphi_\eta(y)|$$

for all $\eta \in V_\delta(\Theta)$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

(v) There exist functions $\phi, \psi : \mathcal{Y} \to (0,\infty)$ such that $\int \phi(y)\nu(dy) < \infty$ and

$$|\varphi_\eta(y)| \leq \phi(y), \qquad |\log|\varphi_\eta(y)|| \leq \psi(y)$$

for all $\eta \in V_\delta(\Theta)$, $y \in \mathcal{Y}$.

Assumption 2.3. *There exists a real number $\gamma \in (0,1)$ such that*

$$\int r_\theta(y,x'|x)\mu(dx') \geq \gamma |\varphi_\eta(y)|$$

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

Assumption 2.4. *$\int \psi(y)\phi(y)\nu(dy) < \infty$.*

Assumption 2.5. *There exists a real number $K \in [1,\infty)$ such that*

$$\int \psi(y) Q(x,dy) \leq K$$

for all $x \in \mathcal{X}$. Moreover, there exist a probability measure $\pi(dx)$ on $\mathcal{X}$ and a real number $\rho \in (0,1)$ such that

$$|P^n(x,B) - \pi(B)| \leq K\rho^n \tag{3}$$

for all $x \in \mathcal{X}$, $n \geq 0$ and any Borel-set $B \subseteq \mathcal{X}$.

Assumption 2.1 is related to the stability of the hidden Markov model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$ and its optimal filter. This assumption ensures that the Markov chain $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$ is geometrically ergodic (see Lemma 5.4) and that the optimal filter for the model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$ forgets initial conditions at an exponential rate (see Lemma 6.2). In this or similar form, Assumption 2.1 is an ingredient of a number of asymptotic results on optimal filtering and maximum likelihood estimation in hidden Markov models (see [4], [5], [16], [17].

Assumption 2.2 is an assumption on the parameterization of the model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$. It requires the transition kernel and density of the chain $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$ to be real-analytic in parameter θ. Together with Assumption 2.1, Assumption 2.2 ensures that an analytic continuation of this kernel exists and is geometrically ergodic (see Lemma 5.4).

Assumption 2.3 is also related to the parameterization of the model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$. This assumption ensures that the ratio

$$\frac{r_\theta(y,x'|x)}{\int r_\theta(y,x''|x)\mu(dx'')}$$

is uniformly bounded in θ, x, x'. Together with Assumptions 2.1 and 2.2, Assumption 2.3 ensures that an analytic continuation of the optimal filter for the model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$ exists and forgets initial conditions at an exponential rate (see Lemma 6.6).

Assumption 2.4 corresponds to the bounding functions $\phi(y)$, $\psi(y)$ (introduced in Assumption 2.2). It requires the product of these functions to be integrable with respect to the measure $\nu(dy)$. Together with Assumption 2.2, Assumption 2.4 ensures that the entropy $h_n(\theta,\lambda)$ (defined in (1)) exists and has an analytic continuation in θ (see Lemma 7.2).

Assumption 2.5 ensures that the Markov chain $\{(X_n, Y_n)\}_{n\geq 0}$ is geometrically ergodic (see Lemma 5.1). Together with Assumption 2.5, Assumption 2.2 also ensures that the log-likelihood $l_n(\theta,\lambda)$ (defined in (2)) exists and admits an analytic continuation.

The following two theorems are the main results of the paper.

Theorem 2.1. *Let Assumptions 2.1 – 2.3 and 2.5 hold. Then, there exists a function $l : \Theta \to \mathbb{R}$ such that $l(\theta)$ is real-analytic for each $\theta \in \Theta$ and $l(\theta) = \lim_{n\to\infty} l_n(\theta,\lambda)$ for all $\theta \in \Theta$, $\lambda \in \mathcal{P}(\mathcal{X})$.*

Theorem 2.2. *Let Assumptions 2.1 – 2.4 hold. Then, there exists a function $h : \Theta \to \mathbb{R}$ such that $h(\theta)$ is real-analytic for each $\theta \in \Theta$ and $h(\theta) = \lim_{n\to\infty} h_n(\theta,\lambda)$ for all $\theta \in \Theta$, $\lambda \in \mathcal{P}(\mathcal{X})$.*

Remark. *As Θ can be represented as a union of open balls, it is sufficient to show Theorems 2.1 and 2.2 for the case where Θ is convex and bounded. Therefore, throughout the analysis carried out in Sections V – VIII, we assume that Θ is a bounded open convex set.*

Theorems 2.1 and 2.2 are proved in Section VII. According to these theorems, for all $\theta \in \Theta$, $\lambda \in \mathcal{P}(\mathcal{X})$, rates $\lim_{n\to\infty} h_n(\theta,\lambda)$ and $\lim_{n\to\infty} l_n(\theta,\lambda)$ are well-defined. Moreover, for each $\theta \in \Theta$, the rates $\lim_{n\to\infty} h_n(\theta,\lambda)$ and $\lim_{n\to\infty} l_n(\theta,\lambda)$ are independent of λ and real-analytic in θ.

The analytical properties of the entropy rates of hidden Markov models have already been extensively studied in several papers [10] – [14], [19], [20], [22], [23]. However, the results presented therein apply exclusively to models with finite state-spaces. To the best of our knowledge, Theorems 2.1

and 2.2 are the first results on the analyticity of the entropy rates of continuous-state hidden Markov models. These theorems also generalize the existing results on the analyticity of the entropy rates of finite-state hidden Markov models.More specifically, [12] can be considered as the strongest existing result of this kind. Theorem 2.2 includes, as a particular case, the results of [12] and simplifies the conditions under which these results hold (for details, see Appendix 2). Theorems 2.1 and 2.2 are relevant for several theoretically and practically important problems arising in statistical inference and system identification. In [26], we rely on these theorems to analyze recursive maximum likelihood estimation in non-linear state-space models. The same theorems can also be used to study the higher-order statistical asymptotics for maximum likelihood estimation in time-series models (for details on such asymptotics, see [27]).

III. Example: Mixture of Densities

In this section, the main results are applied to the case when $p_\theta(x'|x)$ and $q_\theta(y|x)$ are mixtures of probability densities, i.e.

$$p_\theta(x'|x) = \sum_{i=1}^{N_x} a_\theta^i(x) v_i(x'), \qquad q_\theta(y|x) = \sum_{j=1}^{N_y} b_\theta^j(x) w_j(y) \tag{4}$$

for $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ (Θ, $\mathcal{X}$, $\mathcal{Y}$ have the same meaning as in the previous section). Here, $N_x > 1$ and $N_y > 1$ are integers. $\{v_i(x)\}_{1 \leq i \leq N_x}$ and $\{w_j(y)\}_{1 \leq j \leq N_y}$ are functions which map $x \in \mathcal{X}$, $y \in \mathcal{Y}$ to $[0, \infty)$ and satisfy

$$\int v_i(x) \mu(dx) = \int w_j(y) \nu(dy) = 1$$

for each $1 \leq i \leq N_x$, $1 \leq j \leq N_y$ ($\mu(dx)$, $\nu(dy)$ have the same meaning as in the previous section). $\{a_\theta^i(x)\}_{1 \leq i \leq N_x}$ and $\{b_\theta^j(x)\}_{1 \leq j \leq N_y}$ are functions which map $\theta \in \Theta$, $x \in \mathcal{X}$ to $[0, \infty)$ and satisfy

$$\sum_{i=1}^{N_x} a_\theta^i(x) = \sum_{j=1}^{N_y} b_\theta^j(x) = 1.$$

Under these conditions, $v_i(x)$ and $w_j(y)$ are probability densities on $\mathcal{X}$ and $\mathcal{Y}$ (respectively), while $a_\theta^i(x)$ and $b_\theta^j(x)$ are probability masses in i and j (respectively). Hence, in x', y, $p_\theta(x'|x)$ and $q_\theta(y|x)$ are mixture of probability densities. $v_i(x)$ and $w_j(y)$ are the components of these mixtures, while $a_\theta^i(x)$ and $b_\theta^j(x)$ are the corresponding weights.

The entropy rates of hidden Markov model specified in (4) are studied under the following assumptions.

Assumption 3.1. *$\mathcal{X}$ is a compact set.*

Assumption 3.2. *$a_\theta^i(x) > 0$ and $b_\theta^j(x) > 0$ for all $\theta \in \Theta$, $x \in \mathcal{X}$, $1 \leq i \leq N_x$, $1 \leq j \leq N_y$. Moreover, $a_\theta^i(x)$ and $b_\theta^j(x)$ are real-analytic in (θ, x) for each $\theta \in \Theta$, $x \in \mathcal{X}$, $1 \leq i \leq N_x$, $1 \leq j \leq N_y$.*

Assumption 3.3. *There exists a real number $\varepsilon \in (0, 1)$ such that $\varepsilon \leq v_i(x) \leq 1/\varepsilon$ for all $x \in \mathcal{X}$, $1 \leq i \leq N_x$.*

Assumption 3.4. *$\int |\log w_k(y)| w_j(y) \nu(dy) < \infty$ for each $1 \leq j, k \leq N_y$.*

Assumption 3.5. *There exists a real number $K \in [1, \infty)$ such that*

$$\int |\log w_k(y)| Q(x, dy) < \infty$$

for all $x \in \mathcal{X}$, $1 \leq k \leq N_y$. Moreover, there exist a probability measure $\pi(dx)$ on $\mathcal{X}$ and a real number $\rho \in (0, 1)$ such that (3) holds for all $x \in \mathcal{X}$, $n \geq 0$ and any Borel-measurable set $B \subseteq \mathcal{X}$.

Assumptions 3.1 – 3.5 cover several classes of hidden Markov models met in practice. These assumptions indeed hold if $q_\theta(y|x)$ is a mixture of Gamma, Gaussian, Pareto and logistic distributions, and if $p_\theta(x'|x)$ is a mixture of the same distributions truncated to a compact domain.

Using Theorem 2.1 and Theorem 2.2, we obtain the following results.

Corollary 3.1. *Let Assumptions 3.1 – 3.3 and 3.5 hold. Then, all conclusions of Theorem 2.1 are true.*

Corollary 3.2. *Let Assumptions 3.1 – 3.4 hold. Then, all conclusions of Theorem 2.2 are true.*

Corollaries 3.1 and 3.2 are proved in Section VIII.

IV. Example: Non-Linear State-Space Models

In this section, the main results are used to study the entropy rates of non-linear state-space models. We consider the following parameterized state-space model:

$$X_{n+1}^{\theta,\lambda} = A_\theta(X_n^{\theta,\lambda}) + B_\theta(X_n^{\theta,\lambda}) V_n, \tag{5}$$
$$Y_n^{\theta,\lambda} = C_\theta(X_n^{\theta,\lambda}) + D_\theta(X_n^{\theta,\lambda}) W_n, \qquad n \geq 0. \tag{6}$$

Here $\theta \in \Theta$, $\lambda \in \mathcal{P}(\mathcal{X})$ are the parameters indexing the state-space model (5), (6) (Θ, $\mathcal{P}(\mathcal{X})$ have the same meaning as in Section II). $A_\theta(x)$ and $B_\theta(x)$ are functions which map $\theta \in \Theta$, $x \in \mathbb{R}^{d_x}$ (respectively) to $\mathbb{R}^{d_x}$ and $\mathbb{R}^{d_x \times d_x}$ (d_x has the same meaning as in Section II). $C_\theta(x)$ and $D_\theta(x)$ are functions which map $\theta \in \Theta$, $x \in \mathbb{R}^{d_x}$ (respectively) to $\mathbb{R}^{d_y}$ and $\mathbb{R}^{d_y \times d_y}$ (d_y has the same meaning as in Section II). $X_0^{\theta,\lambda}$ is an $\mathbb{R}^{d_x}$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ and distributed according to λ. $\{V_n\}_{n \geq 0}$ are $\mathbb{R}^{d_x}$-valued i.i.d. random variables which are defined on $(\Omega, \mathcal{F}, P)$ and have (marginal) probability density $v(x)$ with respect to the Lebesgue measure. $\{W_n\}_{n \geq 0}$ are $\mathbb{R}^{d_y}$-valued i.i.d. random variables which are defined on $(\Omega, \mathcal{F}, P)$ and have (marginal) probability density $w(y)$ with respect to the Lebesgue measure. We also assume that $X_0^{\theta,\lambda}$, $\{V_n\}_{n \geq 0}$ and $\{W_n\}_{n \geq 0}$ are (jointly) independent.

We also use the following notations. For $\theta \in \Theta$, $x, x' \in \mathbb{R}^{d_x}$, $y \in \mathbb{R}^{d_y}$, $\tilde{p}_\theta(x'|x)$ and $\tilde{q}_\theta(y|x)$ are the functions defined by

$$\tilde{p}_\theta(x'|x) = \frac{v\left(B_\theta^{-1}(x)(x' - A_\theta(x))\right)}{|\det B_\theta(x)|},$$
$$\tilde{q}_\theta(y|x) = \frac{w\left(D_\theta^{-1}(x)(y - C_\theta(x))\right)}{|\det D_\theta(x)|}$$

(provided that $B_\theta(x)$, $D_\theta(x)$ are invertible) while $p_\theta(x'|x)$ and $q_\theta(y|x)$ are defined by

$$p_\theta(x'|x) = \frac{v\left(B_\theta^{-1}(x)(x' - A_\theta(x))\right) 1_{\mathcal{X}}(x')}{\int_{\mathcal{X}} v\left(B_\theta^{-1}(x)(x'' - A_\theta(x))\right) dx''}, \tag{7}$$

$$q_\theta(y|x) = \frac{w\left(D_\theta^{-1}(x)(y - C_\theta(x))\right) 1_{\mathcal{Y}}(y)}{\int_{\mathcal{Y}} w\left(D_\theta^{-1}(x)(y' - C_\theta(x))\right) dy'} \tag{8}$$

It is straightforward to show 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 $X_n^{\theta,\lambda} = x$. $p_\theta(x'|x)$ and $q_\theta(y|x)$ can be interpreted as truncations of $\tilde{p}_\theta(x'|x)$ and $\tilde{q}_\theta(y|x)$ to domains $\mathcal{X}$ and $\mathcal{Y}$ (i.e., the hidden Markov model specified in (7), (8) can be viewed as a truncated version of the original model (5), (6). $p_\theta(x'|x)$ and $q_\theta(y|x)$ accurately approximate $\tilde{p}_\theta(x'|x)$ and $\tilde{q}_\theta(y|x)$ when domains $\mathcal{X}$ and $\mathcal{Y}$ are sufficiently large (i.e., when $\mathcal{X}$, $\mathcal{Y}$ contain balls of sufficiently large radius). This kind of approximation is involved (implicitly or explicitly) in any numerical implementation of the optimal filter for state-space model (5), (6) (for details see e.g., [2], [3], [6]).

The entropy rates of the hidden Markov model (7), (8) are studied under the following assumptions.

Assumption 4.1. *$\mathcal{X}$ and $\mathcal{Y}$ are compact sets with non-empty interiors.*

Assumption 4.2. *$v(x) > 0$ and $w(y) > 0$ for all $x \in \mathbb{R}^{d_x}$, $y \in \mathbb{R}^{d_y}$. Moreover, $v(x)$ and $w(y)$ are real-analytic for each $x \in \mathbb{R}^{d_x}$, $y \in \mathbb{R}^{d_y}$.*

Assumption 4.3. *$B_\theta(x)$ and $D_\theta(x)$ are invertible for all $\theta \in \Theta$, $x \in \mathbb{R}^{d_x}$. Moreover, $A_\theta(x)$, $B_\theta(x)$, $C_\theta(x)$ and $D_\theta(x)$ are real-analytic in (θ, x) for each $\theta \in \Theta$, $x \in \mathbb{R}^{d_x}$.*

Assumption 4.4. *There exist a probability measure $\pi(dx)$ on $\mathcal{X}$ and a real number $\rho \in (0,1)$ such that (3) holds for all $x \in \mathcal{X}$, $n \geq 0$ and any Borel-measurable set $B \subseteq \mathcal{X}$.*

Assumptions 4.1 – 4.3 are relevant for several practically important classes of non-linear state-space models. E.g., these assumptions cover stochastic volatility and dynamic probit models and their truncated versions. For other models satisfying (5), (6) and Assumptions 4.1 – 4.3, see [2], [3], [6] and references cited therein.

Using Theorems 2.1 and 2.2, we get the following results.

Corollary 4.1. *Let Assumptions 4.1 – 4.3 and 4.4 hold. Then, all conclusions of Theorem 2.1 are true.*

Corollary 4.2. *Let Assumptions 4.1 – 4.3 hold. Then, all conclusions of Theorem 2.2 are true.*

Corollaries 4.1 and 4.2 are proved in Section VIII.

V. Results Related to Kernels of $\{(X_n, Y_n)\}_{n\geq 0}$ and $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$

In this section, an analytical (complex-valued) continuation of the transition kernel of $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$ is constructed, and its asymptotic properties (geometric ergodicity) are studied. The same properties of the transition kernel of $\{(X_n, Y_n)\}_{n\geq 0}$ are studied, too. Throughout this and later sections, the following notations is used. Let $\mathcal{W}$ be any Borel set in $\mathbb{R}^{d_w}$, where d_w is any positive integer. Then, $\mathcal{B}(\mathcal{W})$ denotes the collection of Borel sets in $\mathcal{W}$. $\mathcal{P}(\mathcal{W})$ is the collection of probability measures on $\mathcal{W}$, while $\mathcal{M}_p(\mathcal{W})$ is the set of positive measures on $\mathcal{W}$. $\mathcal{M}_c(\mathcal{W})$ is the collection of complex measures on $\mathcal{W}$, while $\mathcal{P}_c(\mathcal{W})$ is the set defined by

$$\mathcal{P}_c(\mathcal{W}) = \{\zeta \in \mathcal{M}_c(\mathcal{W}) : \zeta(\mathcal{W}) = 1\}.$$

For $\zeta \in \mathcal{M}_c(\mathcal{W})$, $\|\zeta\|$ denotes the total variation norm of ζ, while $|\zeta|(dw)$ is the total variation of $\zeta(dw)$. For $w \in \mathcal{W}$, $\delta_w(dw')$ is the Dirac measure centered at w (i.e., $\delta_w(B) = I_B(w)$ for $B \in \mathcal{B}(\mathcal{W})$).

We also need to introduce additional notations. $\mathcal{Z}$ is the set defined by $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$. $\hat{s}_\eta(x)$ and $\tilde{r}_\eta(y, x'|x)$ are the functions defined by

$$\hat{s}_\eta(x) = \int\int \hat{r}_\eta(y', x''|x) \nu(dy') \mu(dx''), \tag{9}$$

$$\tilde{r}_\eta(y, x'|x) = \begin{cases} \hat{r}_\eta(y, x'|x)/\hat{s}_\eta(x), & \text{if } \hat{s}_\eta(x) \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

for $\eta \in \mathbb{C}^d$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. $\tilde{\psi}(z)$ is the function defined by

$$\tilde{\psi}(z) = 1 + \psi(y), \tag{11}$$

where $z = (y, x)$. $u_\eta^n(x_{0:n}, y_{1:n})$ is the function defined by

$$u_\eta^n(x_{0:n}, y_{1:n}) = \prod_{k=1}^n \tilde{r}_\eta(y_k, x_k|x_{k-1}), \tag{12}$$

where $x_0, \dots, x_n \in \mathcal{X}$, $y_1, \dots, y_n \in \mathcal{Y}$, $n \geq 1$. $\sigma(dz)$ is the measure defined by

$$\sigma(B) = \int\int I_B(y, x) Q(x, dy) \pi(dx)$$

for $B \in \mathcal{B}(\mathcal{Z})$. $S(z, dz')$, $S_\eta(z, dz')$ are the kernels defined by

$$S(z, B) = \int\int I_B(y', x') Q(x', dy') P(x, dx'), \tag{13}$$

$$S_\eta(z, B) = \int\int I_B(y', x') \tilde{r}_\eta(y', x'|x) \nu(dy') \mu(dx') \tag{14}$$

(notice that $z = (y, x)$). $\{S^n(z, dz')\}_{n\geq 0}$, $\{S_\eta^n(z, dz')\}_{n\geq 0}$ are the kernels recursively defined by $S^0(z, B) = S_\eta^0(z, B) = \delta_z(B)$ and

$$S^{n+1}(z, B) = \int S^n(z', B) S(z, dz'),$$

$$S_\eta^{n+1}(z, B) = \int S_\eta^n(z', B) S_\eta(z, dz').$$

$\{(S_\eta^n \zeta)(dz)\}_{n\geq 0}$ are the measures defined by

$$(S_\eta^n \zeta)(B) = \int S_\eta^n(z, B) \zeta(dz),$$

where $\zeta \in \mathcal{M}_c(\mathcal{Z})$.

Remark. *$S(z, dz')$ and $\sigma(dz)$ are the transition kernel and the invariant distribution of $\{(X_n, Y_n)\}_{n\geq 0}$. When $\theta \in \Theta$, $S_\theta(z, dz')$ boils down to the transition kernel of $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$. Hence, for $\eta \in \mathbb{C}^d$, $S_\eta(z, dz')$ can be* *considered as a complex-valued continuation of the transition kernel of* $\left\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\right\}_{n\geq 0}$. *Kernel* $S_\eta^n(z,dz')$ *admits the representation*

$$\begin{aligned}(S_\eta^n\zeta)(B) = \int\cdots\int\int I_B(y_n,x_n)u_\eta^n(x_{0:n},y_{1:n}) \\ \cdot(\nu\times\mu)(dy_n,dx_n)\cdots(\nu\times\mu)(dy_1,dx_1) \\ \cdot\zeta(dy_0,dx_0). \end{aligned} \tag{15}$$

This representation is used to show that $S_\eta(z,dz')$ *is geometrically ergodic (see Lemma 5.4 and its proof). It is also used to show the analyticity of integral (111) (see Lemma 7.2 and its proof).*

Remark. *Throughout this section and later sections, the following convention is applied. Diacritic* ˜ *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 Assumption 2.5 hold. Then, there exists a real number* $C_1\in[1,\infty)$ *such that*

$$\int\tilde{\psi}(z')S(z,dz')\leq C_1, \qquad |S^n-\sigma|(z,B)\leq C_1\rho^n$$

for all $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$, $n\geq 0$ *(here,* $|S^n-\sigma|(z,dz')$ *denotes the total variation of* $S^n(z,dz')-\sigma(dz')$*, while* ρ *is specified in Assumption 2.5).*

Proof. Let $C_1=2K$ (K is specified in Assumption 2.5). Moreover, let x, y be any elements of $\mathcal{X}$, $\mathcal{Y}$ (respectively), while $z=(y,x)$ (notice that z is any element of $\mathcal{Z}$). Then, we have

$$\begin{aligned}\int\tilde{\psi}(z')S(z,dz') =& \int\int\left(1+\psi(y')\right)Q(x',dy')P(x,dx') \\ \leq& 1+K\leq C_1. \end{aligned}$$

We also have

$$\begin{aligned}&|S^n(z,B)-\sigma(B)| \\ &=\left|\int\int I_B(y',x')Q(x',dy')(P^n-\pi)(x,dx')\right| \\ &\leq\int\int I_B(y',x')Q(x',dy')|P^n-\pi|(x,dx') \\ &\leq 2K\rho^n\leq C_1\rho^n \end{aligned}$$

for $B\in\mathcal{B}(\mathcal{Z})$, $n\geq 0$. □

Lemma 5.2. *Let Assumption 2.2 hold. Then, the following is true:*

(i) $\tilde{r}_\theta(y,x'|x)=r_\theta(y,x'|x)$ *for all* $\theta\in\Theta$, $x,x'\in\mathcal{X}$, $y\in\mathcal{Y}$.

(ii) There exists a real number $\delta_1\in(0,\delta]$ *such that* $\tilde{r}_\eta(y,x'|x)$ *is analytic in* η *and satisfies*

$$\begin{aligned}&|\tilde{r}_\eta(y,x'|x)|\leq 2|\varphi_\eta(y)|, \\ &\int\int\tilde{r}_\eta(y',x''|x)\nu(dy')\mu(dx'')=1 \end{aligned}$$

for all $\eta\in V_{\delta_1}(\Theta)$, $x,x'\in\mathcal{X}$, $y\in\mathcal{Y}$ *(*δ *is specified in Assumption 2.2).*

Remark. *As a direct consequence of Lemma 5.2 (Part (ii)), we have* $S_\eta^n\zeta\in\mathcal{P}_c(\mathcal{Z})$ *(i.e.,* $(S_\eta^n\zeta)(\mathcal{Z})=1$*) for* $\eta\in V_{\delta_1}(\Theta)$, $\zeta\in\mathcal{P}_c(\mathcal{Z})$, $n\geq 1$.

Proof. Due to Assumption 2.2, we have

$$\int\int\phi(y)\nu(dy)\mu(dx)=\|\mu\|\int\phi(y)\nu(dy)<\infty.$$

Then, using Assumption 2.2 and Lemma A1.1 (see Appendix 1), we conclude that $\hat{s}_\eta(x)$ is analytic in η for each $\eta\in V_\delta(\Theta)$, $x\in\mathcal{X}$. Relying on the same arguments, we deduce

$$|\hat{r}_{\eta'}(y,x'|x)-\hat{r}_{\eta''}(y,x'|x)|\leq\frac{d\,\phi(y)\|\eta'-\eta''\|}{\delta} \tag{16}$$

for $\eta',\eta''\in V_\delta(\Theta)$, $x,x'\in\mathcal{X}$, $y\in\mathcal{Y}$ (here, d denotes the dimension of vectors in Θ, $V_\delta(\Theta)$).

Throughout the rest of the proof, the following notations is used. $\tilde{C}$, δ_1 are the real numbers defined by

$$\tilde{C}=\frac{d\|\mu\|}{\delta}\int\phi(y)\nu(dy), \qquad \delta_1=\min\left\{\delta,\frac{1}{2\tilde{C}}\right\}.$$

η, η', η'' are any elements in $V_{\delta_1}(\Theta)$, while θ is any element of Θ satisfying $\|\eta-\theta\|<\delta_1$. x, x' are any elements of $\mathcal{X}$, while y is any element in $\mathcal{Y}$.

Using (16), we conclude

$$\begin{aligned}&\left|\int\int\left(\hat{r}_{\eta'}(y,x'|x)-\hat{r}_{\eta''}(y,x'|x)\right)\nu(dy)\mu(dx')\right| \\ &\leq\int\int|\hat{r}_{\eta'}(y,x'|x)-\hat{r}_{\eta''}(y,x'|x)|\,\nu(dy)\mu(dx') \\ &\leq\frac{d\|\mu\|\|\eta'-\eta''\|}{\delta}\int\phi(y)\nu(dy)=\tilde{C}\|\eta'-\eta''\|. \end{aligned}$$

Consequently, we have

$$\begin{aligned}|\hat{s}_\eta(x)|=&\left|\int\int\hat{r}_\eta(y,x'|x)\nu(dy)\mu(dx')\right| \\ \geq&\int\int\hat{r}_\theta(y,x'|x)\nu(dy)\mu(dx') \\ &-\left|\int\int\left(\hat{r}_\eta(y,x'|x)-\hat{r}_\theta(y,x'|x)\right)\nu(dy)\mu(dx')\right| \\ \geq& 1-\tilde{C}\|\eta-\theta\|\geq\frac{1}{2} \end{aligned}$$

Hence, we get

$$|\hat{s}_\eta(x)|=\left|\int\int\hat{r}_\eta(y,x'|x)\nu(dy)\mu(dx')\right|\geq\frac{1}{2}. \tag{17}$$

Therefore, we have

$$\tilde{r}_\eta(y,x'|x)=\frac{\hat{r}_\eta(y,x'|x)}{\hat{s}_\eta(x)}. \tag{18}$$

As $\hat{s}_\eta(x)$ is analytic in η for each $\eta\in V_{\delta_1}(\Theta)$, we conclude from Assumption 2.2 and (17), (18) that (i), (ii) are true. □

Lemma 5.3. *Let Assumption 2.2 hold. Then, the following is true:*

(i) $u_\eta^n(x_{0:n},y_{1:n})$ *is analytic in* η *for all* $\eta\in V_{\delta_1}(\Theta)$, $x_0,\dots,x_n\in\mathcal{X}$, $y_1,\dots,y_n\in\mathcal{Y}$, $n\geq 1$ *(*δ_1 *is specified in Lemma 5.2).*

(ii) There exists a non-decreasing sequence $\{K_n\}_{n\geq 1}$ in $[1,\infty)$ such that

$$\left|u_\eta^n(x_{0:n},y_{1:n})\right| \leq K_n\left(\prod_{k=1}^n \phi(y_k)\right),$$
$$\left|u_{\eta'}^n(x_{0:n},y_{1:n}) - u_{\eta''}^n(x_{0:n},y_{1:n})\right| \leq K_n\|\eta'-\eta''\|\left(\prod_{k=1}^n \phi(y_k)\right)$$

for all $\eta,\eta',\eta''\in V_{\delta_1}(\Theta)$, $x_0,\dots,x_n\in\mathcal{X}$, $y_1,\dots,y_n\in\mathcal{Y}$, $n\geq 1$.

Proof. Throughout the proof, the following notations is used. $\{K_n\}_{n\geq 1}$ are the real numbers defined by $K_n = 2^n d/\delta_1$ for $n\geq 1$ (here, d denotes the dimension of vectors in Θ, $V_\delta(\Theta)$). η, η', η'' are any elements of $V_{\delta_1}(\Theta)$. $\{x_n\}_{n\geq 0}$, $\{y_n\}_{n\geq 1}$ are any sequences in $\mathcal{X}$, $\mathcal{Y}$ (respectively).

Owing to Lemma 5.2, $u_\eta^n(x_{0:n},y_{1:n})$ is analytic in η for each $\eta\in V_{\delta_1}(\Theta)$. Due to Assumption 2.2 and the same lemma, we have

$$\left|u_\eta^n(x_{0:n},y_{1:n})\right| \leq 2^n\left(\prod_{k=1}^n |\varphi_\eta|(y_k)\right) \leq K_n\left(\prod_{k=1}^n \phi(y_k)\right)$$

for $n\geq 1$. Consequently, Lemma A1.1 (see Appendix 1) yields

$$\begin{aligned}&\left|u_{\eta'}^n(x_{0:n},y_{1:n}) - u_{\eta''}^n(x_{0:n},y_{1:n})\right|\\ &\leq \frac{2^n d\|\eta'-\eta''\|}{\delta_1}\left(\prod_{k=1}^n \phi(y_k)\right) = K_n\|\eta'-\eta''\|\left(\prod_{k=1}^n \phi(y_k)\right)\end{aligned}$$

for $n\geq 1$. □

Lemma 5.4. *Let Assumptions 2.1, 2.2 and 2.4 hold. Then, the following is true:*

(i) There exist real numbers $\delta_2\in(0,\delta_1]$, $C_2\in[1,\infty)$ such that

$$|S_{\eta'} - S_{\eta''}|(z,B) \leq C_2\|\eta'-\eta''\|,$$
$$\int \tilde{\psi}(z')\,|S_\eta|(z,dz') \leq C_2$$

for all $\eta,\eta',\eta''\in V_{\delta_2}(\Theta)$, $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$ (here, $|S_{\eta'}-S_{\eta''}|(z,dz')$ denotes the total variation of $S_{\eta'}(z,dz') - S_{\eta''}(z,dz')$, while δ_1 is specified in Lemma 5.2).

(ii) For each $\eta\in V_{\delta_2}(\Theta)$, there exists a complex measure $\sigma_\eta(dz)$ on $\mathcal{Z}$ such that $\sigma_\eta(B) = \lim_{n\to\infty} S_\eta^n(z,B)$ for all $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$.

(iii) There exists a real number $\gamma_1\in(0,1)$, such that

$$\left|S_\eta^n - \sigma_\eta\right|(z,B) \leq C_2\gamma_1^n$$

for all $\eta\in V_{\delta_2}(\Theta)$, $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$, $n\geq 0$ (here, $\left|S_\eta^n - \sigma_\eta\right|(z,dz')$ stands for the total variation of $S_\eta^n(z,dz') - \sigma_\eta(dz')$).

Proof. Throughout the proof, the following notations is used. $\tilde{C}_1$, $\tilde{C}_2$ are the real numbers defined by

$$\tilde{C}_1 = 2\|\mu\|\int \psi(y)\phi(y)\nu(dy), \quad \tilde{C}_2 = 2\|\mu\|\int \phi(y)\nu(dy),$$

while n_0 is the integer defined as

$$n_0 = \left\lceil \frac{\log 4}{|\log(1-\varepsilon^2)|}\right\rceil$$

(ε, $\phi(y)$, $\psi(y)$, are specified in Assumptions 2.1, 2.3). $\{\tilde{K}_n\}_{n\geq 1}$ are the real numbers defined by $\tilde{K}_n = (1+\tilde{C}_2)^n K_n$ for $n\geq 1$, while δ_2, γ_1, $\tilde{C}_3$, C_2 are the real numbers defined as $\delta_2 = \delta_1/(4\tilde{K}_{n_0})$, $\gamma_1 = 2^{-1/n_0}$ and

$$\tilde{C}_3 = \tilde{K}_1 + \tilde{C}_1 + \tilde{C}_2, \qquad C_2 = 16\tilde{C}_3\gamma_1^{-n_0}$$

(δ_1, K_n are specified in Lemmas 5.2, 5.3). η, η', η'' are any elements in $V_{\delta_2}(\Theta)$, while θ is any element of Θ satisfying $\|\eta-\theta\| < \delta_2$. x, y are any elements of $\mathcal{X}$, $\mathcal{Y}$ (respectively), while $z=(y,x)$. ζ, ζ', ζ'' are any elements of $\mathcal{P}_c(\mathcal{Z})$, while B is any element of $\mathcal{B}(\mathcal{Z})$. $n\geq 1$, $k\geq 0$ are any integers.

Relying on Assumption 2.4 and Lemma 5.2, we deduce

$$\begin{aligned}\int \tilde{\psi}(z')\,|S_\eta|(z,dz') \leq& \int\int (1+\psi(y'))\,|\tilde{r}_\eta(y',x'|x)|\\ &\cdot\nu(dy')\mu(dx')\\ \leq& 2\|\mu\|\int (1+\psi(y'))\,\varphi(y')\nu(dy')\\ =&\tilde{C}_1 + \tilde{C}_2 \leq C_2\end{aligned}$$

(as $\tilde{C}_1+\tilde{C}_2 \leq \tilde{C}_3 \leq C_2$). Moreover, using Lemma 5.3, we conclude

$$\begin{aligned}&\left|(S_{\eta'}^n\zeta)(B) - (S_{\eta''}^n\zeta)(B)\right|\\ &\leq \int\cdots\int\int I_B(y_n,x_n)\left|u_{\eta'}^n(x_{0:n},y_{1:n}) - u_{\eta''}^n(x_{0:n},y_{1:n})\right|\\ &\quad\cdot(\nu\times\mu)(dy_n,dx_n)\cdots(\nu\times\mu)(dy_1,dx_1)|\zeta|(dy_0,dx_0)\\ &\leq K_n\|\mu\|^n\|\zeta\|\|\eta'-\eta''\|\left(\prod_{k=1}^n\int\phi(y_k)\nu(dy_k)\right)\\ &\leq \tilde{K}_n\|\zeta\|\|\eta'-\eta''\|\end{aligned}$$

Therefore, we get

$$\left\|S_{\eta'}^n\zeta - S_{\eta''}^n\zeta\right\| \leq \tilde{K}_n\|\zeta\|\|\eta'-\eta''\|. \tag{19}$$

Hence, we have

$$\begin{aligned}|S_{\eta'} - S_{\eta''}|(z,B) =& |S_{\eta'}\delta_z - S_{\eta''}\delta_z|(B)\\ \leq&\tilde{K}_1\|\delta_z\|\|\eta'-\eta''\|\\ \leq& C_2\|\eta'-\eta''\|\end{aligned}$$

(as $\tilde{K}_1 \leq \tilde{C}_3 \leq C_2$).

Let $\tau_\theta(dz)$ be the measure defined by

$$\tau_\theta(B) = \int\int I_B(y,x)\lambda_\theta(dx|y)\nu(dy).$$

Owing to Assumption 2.1, we have

$$1 = \int\int r_\theta(y,x'|x)\nu(dy)\mu(dx') \leq \frac{1}{\varepsilon}\int\lambda_\theta(\mathcal{X}|y)\nu(dy).$$

Hence, we get

$$\tau_\theta(\mathcal{Z}) = \int\lambda_\theta(\mathcal{X}|y)\nu(dy) \geq \varepsilon.$$

Moreover, due to Assumption 2.1 and Lemma 5.2, we have

$$\begin{aligned} S_\theta(z,B) =& \int\int I_B(y',x') r_\theta(y',x'|x)\nu(dy')\mu(dx') \\ \geq& \varepsilon \int\int I_B(y',x')\lambda_\theta(dx'|y')\nu(dy') = \varepsilon\tau_\theta(B) \end{aligned}$$

Then, standard results in Markov chain theory (see e.g., [18, Theorem 16.0.2]) imply that there exists a probability measure $\sigma_\theta(dz)$ on $\mathcal{Z}$ such that

$$|S_\theta^n(z,B) - \sigma_\theta(B)| \leq (1-\varepsilon\tau_\theta(\mathcal{Z}))^n \leq (1-\varepsilon^2)^n.$$

Consequently, we get

$$\begin{aligned} &|(S_\theta^n\zeta')(B) - (S_\theta^n\zeta'')(B)| \\ &= \left| \int (S_\theta^n - \sigma_\theta)(z,B)(\zeta'-\zeta'')(dz) \right| \\ &\leq \int |S_\theta^n - \sigma_\theta|(z,B)|\zeta'-\zeta''|(dz) \\ &\leq (1-\varepsilon^2)^n \|\zeta'-\zeta''\| \end{aligned}$$

(as $\sigma_\theta(B)(\zeta'(\mathcal{Z}) - \zeta''(\mathcal{Z})) = 0$). Hence, we have

$$\|S_\theta^n\zeta' - S_\theta^n\zeta''\| \leq (1-\varepsilon^2)^n\|\zeta'-\zeta''\|. \tag{20}$$

Since $S_\theta^n(z,dz')$ is an element of $\mathcal{P}(\mathcal{Z})$, we conclude $\|S_\theta^n\zeta\| \leq \|\zeta\|$. Then, owing to (19), we have

$$\begin{aligned} \left\|S_\eta^n\zeta\right\| \leq& \|S_\theta^n\zeta\| + \left\|\left(S_\eta^n - S_\theta^n\right)\zeta\right\| \\ \leq& \left(1 + \tilde{K}_n\|\eta-\theta\|\right)\|\zeta\| \\ \leq& (1+\tilde{K}_n\delta_2)\|\zeta\| \leq 2\|\zeta\| \end{aligned} \tag{21}$$

when $n \leq n_0$ (as $\|\eta-\theta\| < \delta_2$, $\tilde{K}_n\delta_2 \leq \tilde{K}_{n_0}\delta_2 = \delta_1/4 \leq 1/4$). Moreover, due to (19), (20), we have

$$\begin{aligned} \left\|S_\eta^n\zeta' - S_\eta^n\zeta''\right\| \leq& \|S_\theta^n\zeta' - S_\theta^n\zeta''\| + \left\|(S_\eta^n - S_\theta^n)(\zeta'-\zeta'')\right\| \\ \leq& \left((1-\varepsilon^2)^n + \tilde{K}_n\|\eta-\theta\|\right)\|\zeta'-\zeta''\| \\ \leq& \left((1-\varepsilon^2)^n + \frac{1}{4}\right)\|\zeta'-\zeta''\| \end{aligned}$$

when $n \leq n_0$. Setting $n = n_0$, we conclude

$$\left\|S_\eta^{n_0}\zeta' - S_\eta^{n_0}\zeta''\right\| \leq \frac{\|\zeta'-\zeta''\|}{2}$$

(as $(1-\varepsilon^2)^{n_0} \leq 1/4$). Since $S_\eta^n\zeta \in \mathcal{P}_c(\mathcal{Z})$ (see Lemma 5.2 and the remark immediately after its statement), we have

$$\begin{aligned} \left\|S_\eta^{(k+1)n_0}(\zeta'-\zeta'')\right\| =& \left\|S_\eta^{n_0}\left(S_\eta^{kn_0}\zeta' - S_\eta^{kn_0}\zeta''\right)\right\| \\ \leq& \frac{1}{2}\left\|S_\eta^{kn_0}(\zeta'-\zeta'')\right\| \end{aligned} \tag{22}$$

Iterating (22), we get

$$\left\|S_\eta^{kn_0}(\zeta'-\zeta'')\right\| \leq \frac{1}{2^k}\|\zeta'-\zeta''\|. \tag{23}$$

Using (21), (23), we conclude

$$\begin{aligned} \left\|S_\eta^{(k+1)n_0}\zeta - S_\eta^{kn_0}\zeta\right\| =& \left\|S_\eta^{kn_0}\left(S_\eta^{n_0}\zeta - \zeta\right)\right\| \\ \leq& \frac{1}{2^k}\left\|S_\eta^{n_0}\zeta - \zeta\right\| \\ \leq& \frac{1}{2^k}\left(\left\|S_\eta^{n_0}\zeta\right\| + \|\zeta\|\right) \\ \leq& \frac{\|\zeta\|}{2^{k-2}} \end{aligned} \tag{24}$$

Hence, we get

$$\sum_{k=0}^{\infty}\left\|S_\eta^{(k+1)n_0}\zeta - S_\eta^{kn_0}\zeta\right\| \leq \sum_{k=0}^{\infty}\frac{\|\zeta\|}{2^{k-2}} = 8\|\zeta\| < \infty. \tag{25}$$

Let $(S_\eta^\infty\zeta)(dz)$ be the measure defined by

$$(S_\eta^\infty\zeta)(B) = \zeta(B) + \sum_{k=0}^{\infty}\left((S_\eta^{(k+1)n_0}\zeta)(B) - (S_\eta^{kn_0}\zeta)(B)\right).$$

Then, due to (25), $(S_\eta^\infty\zeta)(dz)$ is well-defined and satisfies $S_\eta^\infty\zeta \in \mathcal{P}_c(\mathcal{Z})$. Moreover, owing to (24), (25), we have

$$\begin{aligned} \left\|S_\eta^{kn_0}\zeta - S_\eta^\infty\zeta\right\| =& \left\|\sum_{j=k}^{\infty}\left(S_\eta^{(j+1)n_0}\zeta - S_\eta^{jn_0}\zeta\right)\right\| \\ \leq& \sum_{j=k}^{\infty}\left\|S_\eta^{(j+1)n_0}\zeta - S_\eta^{jn_0}\zeta\right\| \\ \leq& \sum_{j=k}^{\infty}\frac{\|\zeta\|}{2^{j-2}} = \frac{\|\zeta\|}{2^{k-3}} \end{aligned} \tag{26}$$

Combining this with (23), we get

$$\begin{aligned} \left\|S_\eta^\infty\zeta' - S_\eta^\infty\zeta''\right\| \leq& \left\|S_\eta^{kn_0}\zeta' - S_\eta^\infty\zeta'\right\| + \left\|S_\eta^{kn_0}\zeta'' - S_\eta^\infty\zeta''\right\| \\ &+ \left\|S_\eta^{kn_0}\zeta' - S_\eta^{kn_0}\zeta''\right\| \\ \leq& \frac{\|\zeta'\| + \|\zeta''\| + \|\zeta'-\zeta''\|}{2^{k-3}} \end{aligned}$$

Therefore, $S_\eta^\infty\zeta' = S_\eta^\infty\zeta''$ for any $\zeta',\zeta'' \in \mathcal{P}_c(\mathcal{Z})$. Consequently, there exists $\sigma_\eta \in \mathcal{P}_c(\mathcal{Z})$ such that $S_\eta^\infty\zeta = \sigma_\eta$ for any $\zeta \in \mathcal{P}_c(\mathcal{Z})$. Hence, $S_\eta^\infty(S_\eta^n\zeta) = \sigma_\eta$ (notice that $S_\eta^n\zeta \in \mathcal{P}_c(\mathcal{Z})$). Then, (21), (26) imply

$$\begin{aligned} \left\|S_\eta^n\zeta - \sigma_\eta\right\| =& \left\|S_\eta^{kn_0}(S_\eta^{n-kn_0}\zeta) - S_\eta^\infty(S_\eta^{n-kn_0}\zeta)\right\| \\ \leq& \frac{1}{2^{k-3}}\left\|S_\eta^{n-kn_0}\zeta\right\| \\ \leq& \frac{\|\zeta\|}{2^{k-4}} \leq C_2\gamma_1^n\|\zeta\| \end{aligned} \tag{27}$$

when $(k+1)n_0 \geq n > kn_0$ (notice that $2^{-(k-4)} = 16\gamma_1^{kn_0} \leq (16\gamma_1^{-n_0})\gamma_1^n \leq C_2\gamma_1^n$). Thus, we get

$$\begin{aligned} \left|S_\eta^n - \sigma_\eta\right|(z,B) =& \left|S_\eta^n\delta_z - \sigma_\eta\right|(B) \\ \leq& \left\|S_\eta^n\delta_z - \sigma_\eta\right\| \\ \leq& C_2\gamma_1^n\|\delta_z\| = C_2\gamma_1^n \end{aligned}$$

(set $k = \lfloor (n-m)/n_0 \rfloor$ in (27)). □

VI. Results Related to Optimal Filter

In this section, an analytic (complex-valued) continuation of the optimal filter is constructed, and its asymptotic properties (exponential forgetting) are studied. Here, we rely on the following notations. $\mathcal{B}(\mathcal{X})$, $\mathcal{P}(\mathcal{X})$, $\mathcal{M}_p(\mathcal{X})$ and $\mathcal{M}_c(\mathcal{X})$ have been defined at the beginning of Section V. For $x \in \mathcal{X}$, $\xi \in \mathcal{M}_c(\mathcal{X})$, $\|\xi\|$, $|\xi|(dx')$ and $\delta_x(dx')$ are the norm and measures specified at the beginning of Section V, too. For $\gamma \in (0,1)$, $V_\gamma(\mathcal{P}(\mathcal{X}))$ is the open γ-vicinity of $\mathcal{P}(\mathcal{X})$, i.e.,

$$V_\gamma(\mathcal{P}(\mathcal{X})) = \{\xi \in \mathcal{M}_c(\mathcal{X}) : \exists\lambda \in \mathcal{P}(\mathcal{X}), \|\xi - \lambda\| < \gamma\}.$$

$R_{\eta,y}(dx|\xi)$ is the measure defined by

$$R_{\eta,y}(B|\xi) = \int\int I_B(x')\tilde{r}_\eta(y,x'|x)\mu(dx')\xi(dx) \tag{28}$$

for $\eta \in \mathbb{C}^d$, $\xi \in \mathcal{M}_c(\mathcal{X})$, $B \in \mathcal{B}(\mathcal{X})$, $y \in \mathcal{Y}$ ($\tilde{r}_\eta(y,x'|x)$ is specified in Lemma 5.2). $\Phi_{\eta,y}(\xi)$ is the function defined by

$$\Phi_{\eta,y}(\xi) = \begin{cases} \log R_{\eta,y}(\mathcal{X}|\xi), & \text{if } R_{\eta,y}(\mathcal{X}|\xi) \neq 0 \\ 0, & \text{otherwise} \end{cases}. \tag{29}$$

$v_{\eta,\boldsymbol{y}}^{m:n}(x_{m:n})$ and $\varphi_{\eta,\boldsymbol{y}}^{m:n}$ are the functions defined by

$$v_{\eta,\boldsymbol{y}}^{m:n}(x_{m:n}) = \prod_{k=m+1}^{n} \tilde{r}_\eta(y_k, x_k|x_{k-1}), \tag{30}$$

$$\varphi_{\eta,\boldsymbol{y}}^{m:n} = \prod_{k=m+1}^{n} \varphi_\eta(y_k), \tag{31}$$

$r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)$ is the function defined by

$$r_{\eta,\boldsymbol{y}}^{m:n}(x'|x) = \int\int\cdots\int\int v_{\eta,\boldsymbol{y}}^{m:n}(x_{m:n}) \cdot \delta_{x'}(dx_n)\mu(dx_{n-1})\cdots\mu(dx_{m+1})\delta_x(dx_m), \tag{32}$$

where $x, x' \in \mathcal{X}$. $R_{\eta,\boldsymbol{y}}^{m:m}(dx|\xi)$ and $R_{\eta,\boldsymbol{y}}^{m:n}(dx|\xi)$ are the measures defined by $R_{\eta,\boldsymbol{y}}^{m:m}(B|\xi) = \xi(B)$ and

$$R_{\eta,\boldsymbol{y}}^{m:n}(B|\xi) = \int\int I_B(x')r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)\mu(dx')\xi(dx). \tag{33}$$

$f_{\eta,\boldsymbol{y}}^{m:n}(x|\xi)$, $g_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi)$, $h_{\eta,\boldsymbol{y}}^{m:n}(x|x',\xi)$ are the functions defined by

$$g_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi) = \begin{cases} r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)/R_{\eta,\boldsymbol{y}}^{m:n}(\mathcal{X}|\xi), & \text{if } R_{\eta,\boldsymbol{y}}^{m:n}(\mathcal{X}|\xi) \neq 0 \\ 0, & \text{otherwise} \end{cases}, \tag{34}$$

$$f_{\eta,\boldsymbol{y}}^{m:n}(x|\xi) = \int g_{\eta,\boldsymbol{y}}^{m:n}(x|x'',\xi)\xi(dx''), \tag{35}$$

$$h_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi) = -f_{\eta,\boldsymbol{y}}^{m:n}(x'|\xi)\int g_{\eta,\boldsymbol{y}}^{m:n}(x''|x,\xi)\mu(dx'') + g_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi). \tag{36}$$

$F_{\eta,\boldsymbol{y}}^{m:m}(dx|\xi)$ and $F_{\eta,\boldsymbol{y}}^{m:n}(dx|\xi)$ are the measures defined by $F_{\eta,\boldsymbol{y}}^{m:m}(B|\xi) = \xi(B)$ and

$$F_{\eta,\boldsymbol{y}}^{m:n}(B|\xi) = \int I_B(x)f_{\eta,\boldsymbol{y}}^{m:n}(x|\xi)\mu(dx). \tag{37}$$

Throughout this and later sections, measures $R_{\eta,\boldsymbol{y}}^{m:n}(dx|\xi)$, $F_{\eta,\boldsymbol{y}}^{m:n}(dx|\xi)$ are also denoted by $R_{\eta,\boldsymbol{y}}^{m:n}(\xi)$, $F_{\eta,\boldsymbol{y}}^{m:n}(\xi)$ (shorthand notations), while $\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\rangle$, $\langle F_{\eta,\boldsymbol{y}}^{m:n}(\xi)\rangle$ are defined by

$$\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle = R_{\eta,\boldsymbol{y}}^{m:n}(\mathcal{X}|\xi), \quad \left\langle F_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle = F_{\eta,\boldsymbol{y}}^{m:n}(\mathcal{X}|\xi). \tag{38}$$

Remark. *When $\theta \in \Theta$, $\lambda \in \mathcal{P}(\mathcal{X})$, $F_{\theta,\boldsymbol{y}}^{m:n}(\lambda)$ is the optimal filter for the model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$, i.e.,*

$$F_{\theta,\boldsymbol{y}}^{1:n}(B|\lambda) = P\left(X_n^{\theta,\lambda} \in B \middle| Y_{1:n}^{\theta,\lambda} = y_{1:n}\right).$$

Hence, for $\eta \in \mathbb{C}^d$, $\xi \in \mathcal{M}_c(\mathcal{X})$, $F_{\eta,\boldsymbol{y}}^{m:n}(\xi)$ can be considered as a complex-valued continuation of the optimal filter. Consequently, $f_{\theta,\boldsymbol{y}}^{m:n}(x|\xi)$ can be viewed as a complex-valued continuation of the optimal filtering density. $h_{\theta,\boldsymbol{y}}^{m:n}(x'|x,\xi)$ can be described as the Gateaux derivative of $f_{\theta,\boldsymbol{y}}^{m:n}(x|\xi)$ with respect to ξ (see (74) – (76)). $h_{\theta,\boldsymbol{y}}^{m:n}(x'|x,\xi)$ is used to show that $F_{\eta,\boldsymbol{y}}^{m:n}(\xi)$ forgets initial condition ξ at an exponential rate (see Lemmas 6.5, 6.6 and their proofs).

Lemma 6.1. *Let η, ξ be any elements of $\mathbb{C}^d$, $\mathcal{M}_s(\mathcal{X})$ (respectively), while $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ is any sequence in $\mathcal{Y}$. Moreover, let n, m, k be any integers satisfying $n \geq k \geq m$. Then, the following is true:*

(i) $R_{\eta,\boldsymbol{y}}^{m:n}(\xi) = R_{\eta,\boldsymbol{y}}^{k:n}\left(R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)$.

(ii) $\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle = \left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle \left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle$ if $\left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle \neq 0$.

(iii) $F_{\eta,\boldsymbol{y}}^{m:n}(\xi) = F_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)$ if $\left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle \neq 0$ and $\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle \neq 0$.

Proof. (i) When $k = m$ or $k = n$, (i) is trivially satisfied. In what follows in this part of the proof, we assume $n > k > m$.

Owing to (28), we have

$$v_{\eta,\boldsymbol{y}}^{m:n}(x_{m:n}) = v_{\eta,\boldsymbol{y}}^{k:n}(x_{k:n})v_{\eta,\boldsymbol{y}}^{m:k}(x_{m:k})$$

for $x_m, \dots, x_n \in \mathcal{X}$. Combining this with (32), it is easy to show

$$r_{\eta,\boldsymbol{y}}^{m:n}(x'|x) = \int r_{\eta,\boldsymbol{y}}^{k:n}(x'|x'')r_{\eta,\boldsymbol{y}}^{m:k}(x''|x)\mu(dx'')$$

for $x, x' \in \mathcal{X}$. Then, using (33), we conclude

$$\begin{aligned} R_{\eta,\boldsymbol{y}}^{m:n}(B|\xi) &= \int\int\int I_B(x')r_{\eta,\boldsymbol{y}}^{k:n}(x'|x'')r_{\eta,\boldsymbol{y}}^{m:k}(x''|x) \cdot \mu(dx'')\mu(dx')\xi(dx) \\ &= \int\int I_B(x')r_{\eta,\boldsymbol{y}}^{k:n}(x'|x'')R_{\eta,\boldsymbol{y}}^{m:k}(dx''|\xi)\mu(dx') \\ &= R_{\eta,\boldsymbol{y}}^{k:n}\left(B|R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right) \end{aligned}$$

for $B \in \mathcal{B}(\mathcal{X})$. Hence, (i) holds when $n > k > m$.

(ii) We assume $\left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle \neq 0$ (i.e., $R_{\eta,\boldsymbol{y}}^{m:k}(\mathcal{X}|\xi) \neq 0$). Then, using (34), (35), (37), we conclude

$$F_{\eta,\boldsymbol{y}}^{m:k}(\xi) = \frac{R_{\eta,\boldsymbol{y}}^{m:k}(\xi)}{\left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle}. \tag{39}$$

Consequently, we have

$$\left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle = \frac{\left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle}{\left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle}$$

(as $\left\langle R_{\eta,\boldsymbol{y}}^{k:n}(\xi)\right\rangle$ is linear in ξ). Combining this with (i), we get

$$\begin{aligned} \left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle &= \left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle \\ &= \left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle \left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle. \end{aligned} \tag{40}$$

Thus, (ii) is true.

(iii) We assume $\left\langle R_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right\rangle \neq 0$, $\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle \neq 0$. Therefore, (ii) implies $\left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle \neq 0$. Then, using the same arguments as in (ii), we deduce

$$F_{\eta,\boldsymbol{y}}^{m:n}(\xi) = \frac{R_{\eta,\boldsymbol{y}}^{m:n}(\xi)}{\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi)\right\rangle},$$

$$F_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right) = \frac{R_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)}{\left\langle R_{\eta,\boldsymbol{y}}^{k:n}\left(F_{\eta,\boldsymbol{y}}^{m:k}(\xi)\right)\right\rangle}.$$

Combining this with (i) and (39), (40), we get

$$\begin{aligned}F^{m:n}_{\eta,\boldsymbol{y}}(\xi)=&\frac{R^{k:n}_{\eta,\boldsymbol{y}}\left(R^{m:k}_{\eta,\boldsymbol{y}}(\xi)\right)}{\left\langle R^{k:n}_{\eta,\boldsymbol{y}}\left(F^{m:k}_{\eta,\boldsymbol{y}}(\xi)\right)\right\rangle\left\langle R^{m:k}_{\eta,\boldsymbol{y}}(\xi)\right\rangle}\\=&\frac{R^{k:n}_{\eta,\boldsymbol{y}}\left(F^{m:k}_{\eta,\boldsymbol{y}}(\xi)\right)}{\left\langle R^{k:n}_{\eta,\boldsymbol{y}}\left(F^{m:k}_{\eta,\boldsymbol{y}}(\xi)\right)\right\rangle}=F^{k:n}_{\eta,\boldsymbol{y}}\left(F^{m:k}_{\eta,\boldsymbol{y}}(\xi)\right)\end{aligned}$$

(use again that $R^{k:n}_{\eta,\boldsymbol{y}}(\xi)$ is linear in ξ). Hence, (iii) holds. □

Lemma 6.2. *Let Assumption 2.1 hold. Then, there exist real numbers $\delta_3\in(0,\delta_1]$, $\gamma_2\in(0,1)$, $C_3\in[1,\infty)$ such that*

$$\left\|F^{m:n}_{\theta,\boldsymbol{y}}(\lambda')-F^{m:n}_{\theta,\boldsymbol{y}}(\lambda'')\right\|\leq C_3\gamma_2^{n-m}\left\|\lambda'-\lambda''\right\|$$

for all $\theta\in\Theta$, $\lambda',\lambda''\in V_{\delta_3}(\mathcal{P}(\mathcal{X}))\cap\mathcal{M}_p(\mathcal{X})$, $n\geq m\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$ (δ_1 is specified in Lemma 5.2).

Proof. Due to [17, Proposition 4.1, Corollary 4.2] (or [24, Theorem 3.1]) and Lemma 5.2, there exist real numbers $\gamma_2\in(0,1)$, $C_3\in[1,\infty)$ such that

$$\left\|F^{m:n}_{\theta,\boldsymbol{y}}(\lambda')-F^{m:n}_{\theta,\boldsymbol{y}}(\lambda'')\right\|\leq\frac{C_3\gamma_2^{n-m}}{4}\left\|\frac{\lambda'}{\|\lambda'\|}-\frac{\lambda''}{\|\lambda''\|}\right\| \tag{41}$$

for all $\theta\in\Theta$, $\lambda',\lambda''\in\mathcal{M}_p(\mathcal{X})$, $n\geq m\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$ (notice that $F^{m:n}_{\theta,\boldsymbol{y}}(\lambda)=F^{m:n}_{\theta,\boldsymbol{y}}(\lambda/\|\lambda\|)$ for each $\lambda\in\mathcal{M}_p(\mathcal{X})$).

Let $\delta_3=\min\{1/2,\delta_1\}$, while $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ is any sequence in $\mathcal{Y}$. Then, we have $\|\lambda\|\geq 1-\delta_3\geq 1/2$ for $\lambda\in V_{\delta_3}(\mathcal{P}(\mathcal{X}))\cap\mathcal{M}_p(\mathcal{X})$. Consequently, (41) implies

$$\begin{aligned}&\left\|F^{m:n}_{\theta,\boldsymbol{y}}(\lambda')-F^{m:n}_{\theta,\boldsymbol{y}}(\lambda'')\right\|\\&\leq\frac{C_3\gamma_2^{n-m}}{4}\left\|\frac{\lambda'-\lambda''}{\|\lambda'\|}-\frac{\lambda''(\|\lambda'\|-\|\lambda''\|)}{\|\lambda'\|\|\lambda''\|}\right\|\\&\leq\frac{C_3\gamma_2^{n-m}\|\lambda'-\lambda''\|}{2\|\lambda'\|}\\&\leq C_3\gamma_2^{n-m}\|\lambda'-\lambda''\|\end{aligned}$$

for $\theta\in\Theta$, $\lambda',\lambda''\in V_{\delta_3}(\mathcal{P}(\mathcal{X}))\cap\mathcal{M}_p(\mathcal{X})$, $n\geq m\geq 0$ (notice that $2\|\lambda'\|\geq 1$, $\|\lambda''/\|\lambda''\|\|=1$, $|\|\lambda'\|-\|\lambda''\||\leq\|\lambda'-\lambda''\|$). □

Lemma 6.3. *Let Assumptions 2.2 and 2.3 hold. Then, the following is true:*
(i) $\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle$ is analytic in η for all $\eta\in V_{\delta_1}(\Theta)$, $\xi\in V_{\delta_1}(\mathcal{P}(\mathcal{X}))$, $n\geq m\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$ (δ_1 is specified in Lemma 5.2).
(ii) There exists a non-decreasing sequence $\{L_n\}_{n\geq 1}$ in $[1,\infty)$ such that

$$\begin{aligned}&\left|\frac{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle}{\varphi^{m:n}_{\eta,\boldsymbol{y}}}\right|\leq L_{n-m},\\&\left|\frac{\left\langle R^{m:n}_{\eta',\boldsymbol{y}}(\xi')\right\rangle}{\varphi^{m:n}_{\eta',\boldsymbol{y}}}-\frac{\left\langle R^{m:n}_{\eta'',\boldsymbol{y}}(\xi'')\right\rangle}{\varphi^{m:n}_{\eta'',\boldsymbol{y}}}\right|\\&\leq L_{n-m}\left(\|\eta'-\eta''\|+\|\xi'-\xi''\|\right)\end{aligned}$$

for all $\eta,\eta',\eta''\in V_{\delta_1}(\Theta)$, $\xi,\xi',\xi''\in V_{\delta_1}(\mathcal{P}(\mathcal{X}))$, $n>m\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$.
(iii) There exists a non-increasing sequence $\{\alpha_n\}_{n\geq 1}$ in $(0,\delta_1]$ such that

$$\frac{\operatorname{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\right\}}{|\varphi^{m:n}_{\theta,\boldsymbol{y}}|}\geq\frac{1}{L_{n-m}}$$

for all $\eta\in V_{\alpha_{n-m}}(\Theta)$, $\xi\in V_{\alpha_{n-m}}(\mathcal{P}(\mathcal{X}))$, $n>m\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$.
(iv) There exists a non-decreasing sequence $\{M_n\}_{n\geq 1}$ in $[1,\infty)$ such that

$$\begin{aligned}&\max\left\{\left|f^{m:n}_{\eta,\boldsymbol{y}}(x|\xi)\right|,\left|h^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi)\right|\right\}\leq M_{n-m},\\&\left|f^{m:n}_{\eta',\boldsymbol{y}}(x|\xi')-f^{m:n}_{\eta'',\boldsymbol{y}}(x|\xi'')\right|\\&\leq M_{n-m}(\|\eta'-\eta''\|+\|\xi'-\xi''\|)\\&\left|h^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi')-h^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right|\\&\leq M_{n-m}(\|\eta'-\eta''\|+\|\xi'-\xi''\|)\end{aligned}$$

for all $\eta,\eta',\eta''\in V_{\alpha_{n-m}}(\Theta)$, $\xi,\xi',\xi''\in V_{\alpha_{n-m}}(\mathcal{P}(\mathcal{X}))$, $x,x'\in\mathcal{X}$, $n>m\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$.

Proof. (i) and (ii) Throughout these parts of the proof, the following notations is used. $\{\tilde{L}_l\}_{l\geq 1}$, $\{L_l\}_{l\geq 1}$ are the real numbers defined by

$$\tilde{L}_l=\frac{2^{l+1}d}{\delta_1}\left(\|\mu\|+\frac{1}{\gamma}\right)^l,\qquad L_l=2\tilde{L}_l^2 \tag{42}$$

for $l\geq 1$ where γ, K_l are specified in Assumption 2.3 and Lemma 5.3. m, n are any integers satisfying $n>m\geq 0$. In what follows in the proof of (i), (ii), both m, n are kept fixed. η, η', η'' are any elements in $V_{\delta_1}(\Theta)$. ξ, ξ', ξ'' are any elements of $V_{\delta_1}(\mathcal{P}(\mathcal{X}))$. x, x' are any elements of $\mathcal{X}$, while $\boldsymbol{y}=\{y_n\}_{n\geq 0}$ is any sequence in $\mathcal{Y}$.

Using (32), (33), (38), it is straightforward to verify

$$\begin{aligned}\frac{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle}{\varphi^{m:n}_{\eta,\boldsymbol{y}}}=&\int\cdots\int\int\frac{v^{m:n}_{\eta,\boldsymbol{y}}(x_{m:n})}{\varphi^{m:n}_{\eta,\boldsymbol{y}}}\\&\cdot\mu(dx_n)\cdots\mu(dx_{m+1})\xi(dx_m).\end{aligned} \tag{43}$$

Moreover, Lemma 5.2 yields

$$\left|\frac{v^{m:n}_{\eta,\boldsymbol{y}}(x_{m:n})}{\varphi^{m:n}_{\eta,\boldsymbol{y}}}\right|=\prod_{k=m+1}^{n}\left|\frac{\tilde{r}_\eta(y_k,x_k|x_{k-1})}{\varphi_\eta(y_k)}\right|\leq 2^{n-m}. \tag{44}$$

Then, Assumption 2.2 and Lemma A1.1 (see Appendix 1) imply that $\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle/\varphi^{m:n}_{\eta,\boldsymbol{y}}$ is analytic in η for each $\eta\in V_{\delta_1}(\Theta)$ (notice that due to Assumption 2.2, $v^{m:n}_{\eta,\boldsymbol{y}}(x_{m:n})/\varphi^{m:n}_{\eta,\boldsymbol{y}}$ is analytic in η for each $\eta\in V_{\delta_1}(\Theta)$, $x_m,\dots,x_n\in\mathcal{X}$). Consequently, $\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle$ is analytic in η for each $\eta\in V_{\delta_1}(\Theta)$. Hence, (i) holds.

Owing to (43), (44), we have

$$\begin{aligned}\left|\frac{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle}{\varphi^{m:n}_{\eta,\boldsymbol{y}}}\right|\leq&\int\cdots\int\int\left|\frac{v^{m:n}_{\eta,\boldsymbol{y}}(x_{m:n})}{\varphi^{m:n}_{\eta,\boldsymbol{y}}}\right|\\&\cdot\mu(dx_n)\cdots\mu(dx_{m+1})|\xi|(dx_m)\\\leq&2^{n-m}\|\mu\|^{n-m}\|\xi\|\\\leq&\tilde{L}_{n-m}\leq L_{n-m}\end{aligned} \tag{45}$$

(notice that $\|\xi\| \leq 1 + \delta_1 \leq 2$ due to $\xi \in V_{\delta_1}(\mathcal{P}(\mathcal{X}))$). Using similar arguments, we have

$$\left| \frac{\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi') \right\rangle}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} - \frac{\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi'') \right\rangle}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} \right| = \left| \frac{\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi' - \xi'') \right\rangle}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} \right| \leq 2^{n-m} \|\mu\|^{n-m} \|\xi' - \xi''\| \leq \tilde{L}_{n-m} \|\xi' - \xi''\|. \tag{46}$$

Combining Lemma A1.1 and (45), we also get

$$\left| \frac{\left\langle R_{\eta',\boldsymbol{y}}^{m:n}(\xi) \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} - \frac{\left\langle R_{\eta'',\boldsymbol{y}}^{m:n}(\xi) \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} \right| \leq \frac{2^{n-m} d \|\mu\|^{n-m} \|\xi\| \|\eta' - \eta''\|}{\delta_1} \leq \tilde{L}_{n-m} \|\eta' - \eta''\|$$

(notice that $\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle / \varphi_{\eta,\boldsymbol{y}}^{m:n}$ is analytic in η for each $\eta \in V_{\delta_1}(\Theta)$). Then, we have

$$\begin{aligned} &\left| \frac{\left\langle R_{\eta',\boldsymbol{y}}^{m:n}(\xi') \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} - \frac{\left\langle R_{\eta'',\boldsymbol{y}}^{m:n}(\xi'') \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} \right| \\ \leq &\left| \frac{\left\langle R_{\eta',\boldsymbol{y}}^{m:n}(\xi') \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} - \frac{\left\langle R_{\eta'',\boldsymbol{y}}^{m:n}(\xi') \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} \right| + \left| \frac{\left\langle R_{\eta'',\boldsymbol{y}}^{m:n}(\xi') \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} - \frac{\left\langle R_{\eta'',\boldsymbol{y}}^{m:n}(\xi'') \right\rangle}{\varphi_{\eta'',\boldsymbol{y}}^{m:n}} \right| \\ \leq &\tilde{L}_{n-m} \left(\|\eta' - \eta''\| + \|\xi' - \xi''\| \right). \end{aligned} \tag{47}$$

Using (45), (47), we conclude that (ii) is true.

(iii) and (iv) Throughout these parts of the proof, we use the following notations. $\{\tilde{L}_l\}_{l\geq 1}$ has the same meaning as in (42), while $\{\alpha_l\}_{l\geq 1}$, $\{\tilde{M}_l\}_{l\geq 1}$, $\{M_l\}_{l\geq 1}$ are the numbers defined by

$$\alpha_l = \frac{\delta_1}{4\tilde{L}_l^2}, \qquad \tilde{M}_l = 10\tilde{L}_l^4, \qquad M_l = 5\tilde{M}_l^2(\|\mu\| + 1).$$

m, n are any integers satisfying $n > m \geq 0$. In what follows in the proof of (iii), (iv), both m, n are kept fixed. η, η', η'' are any elements of $V_{\alpha_{n-m}}(\Theta)$, while θ is any element of Θ satisfying $\|\eta - \theta\| < \alpha_{n-m}$. ξ, ξ', ξ'' are any elements of $V_{\alpha_{n-m}}(\mathcal{P}(\mathcal{X}))$, while λ is any element of $\mathcal{P}(\mathcal{X})$ satisfying $\|\xi - \lambda\| < \alpha_{n-m}$. x, x' are any elements of $\mathcal{X}$, while $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ is any sequence in $\mathcal{Y}$.

Using Lemma 5.2 and (32), it is straightforward to verify

$$\left\langle R_{\theta,\boldsymbol{y}}^{m:k+1}(\lambda) \right\rangle = \int \cdots \int \int \left(\int r_\theta(y_{k+1}, x_{k+1} | x_k) \mu(dx_{k+1}) \right) \cdot v_{\theta,\boldsymbol{y}}^{m:k}(x_{m:k}) \mu(dx_k) \cdots \mu(dx_{m+1}) \lambda(dx_m)$$

for $k > m$. Consequently, Assumption 2.3 yields

$$\begin{aligned} \left\langle R_{\theta,\boldsymbol{y}}^{m:k+1}(\lambda) \right\rangle \geq &\gamma |\varphi_\theta(y_{k+1})| \int \cdots \int \int v_{\theta,\boldsymbol{y}}^{m:k}(x_{m:k}) \cdot \mu(dx_k) \cdots \mu(dx_{m+1}) \lambda(dx_m) \\ = &\gamma |\varphi_\theta(y_{k+1})| \left\langle R_{\theta,\boldsymbol{y}}^{m:k}(\lambda) \right\rangle \end{aligned} \tag{48}$$

The same arguments also imply

$$\left\langle R_{\theta,\boldsymbol{y}}^{m:m+1}(\lambda) \right\rangle = \int \left(\int r_\theta(y_{m+1}, x_{m+1} | x_m) \mu(dx_{m+1}) \right) \lambda(dx_m) \geq \gamma |\varphi_\theta(y_{m+1})| \|\lambda\| = \gamma |\varphi_\theta(y_{m+1})|.$$

Then, iterating (48), we get

$$\begin{aligned} \left\langle R_{\theta,\boldsymbol{y}}^{m:k+1}(\lambda) \right\rangle \geq &\gamma^{k-m-1} \left(\prod_{l=m+2}^{k+1} |\varphi_\theta(y_l)| \right) \left\langle R_{\theta,\boldsymbol{y}}^{m:m+1}(\lambda) \right\rangle \\ \geq &\gamma^{k-m} \left(\prod_{l=m+1}^{k+1} |\varphi_\theta(y_l)| \right) = \gamma^{k-m} |\varphi_{\theta,\boldsymbol{y}}^{m:n}| \end{aligned}$$

Hence, we have

$$\frac{\left\langle R_{\theta,\boldsymbol{y}}^{m:n}(\lambda) \right\rangle}{|\varphi_{\theta,\boldsymbol{y}}^{m:n}|} \geq \frac{1}{\tilde{L}_{n-m}}$$

(as $\tilde{L}_{n-m} \geq \gamma^{-(n-m)}$). Combining this with (47), we get

$$\begin{aligned} \frac{\mathrm{Re}\left\{ \left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle \right\}}{|\varphi_{\eta,\boldsymbol{y}}^{m:n}|} \geq &\frac{\left\langle R_{\theta,\boldsymbol{y}}^{m:n}(\lambda) \right\rangle}{|\varphi_{\theta,\boldsymbol{y}}^{m:n}|} - \left| \frac{\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} - \frac{\left\langle R_{\theta,\boldsymbol{y}}^{m:n}(\lambda) \right\rangle}{\varphi_{\theta,\boldsymbol{y}}^{m:n}} \right| \\ \geq &\frac{1}{\tilde{L}_{n-m}} - \tilde{L}_{n-m} \left(\|\eta - \theta\| + \|\xi - \lambda\| \right) \\ \geq &\frac{1}{\tilde{L}_{n-m}} - 2\tilde{L}_{n-m} \alpha_{n-m} \\ \geq &\frac{1}{2\tilde{L}_{n-m}} \geq \frac{1}{2\tilde{L}_{n-m}} \end{aligned} \tag{49}$$

(notice that $\|\eta - \theta\| < \alpha_{n-m}$, $\|\xi - \lambda\| < \alpha_{n-m}$).

Using (32), it is straightforward to verify

$$\frac{r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} = \int \int \cdots \int \int \frac{v_{\eta,\boldsymbol{y}}^{m:n}(x_{m:n})}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} \cdot \delta_{x'}(dx_n) \mu(dx_{n-1}) \cdots \mu(dx_{m+1}) \delta_x(dx_m). \tag{50}$$

Then, Assumption 2.2, Lemma A1.1 and (44) imply that $r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)/\varphi_{\eta,\boldsymbol{y}}^{m:n}$ is analytic in η for each $\eta \in V_{\delta_1}(\Theta)$. Consequently, $r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)$ is analytic in η for all $\eta \in V_{\delta_1}(\Theta)$. Combining this with (34), (49), we conclude that $g_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi)$ is analytic in η for each $\eta \in V_{\delta_1}(\Theta)$ (notice that $\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle$ is non-zero and analytic in η for all $\eta \in V_{\delta_1}(\Theta)$).

Owing to (44), (50), we have

$$\begin{aligned} \left| \frac{r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} \right| \leq &\int \int \cdots \int \int \left| \frac{v_{\eta,\boldsymbol{y}}^{m:n}(x_{m:n})}{\varphi_{\eta,\boldsymbol{y}}^{m:n}} \right| \cdot \delta_{x'}(dx_n) \mu(dx_{n-1}) \cdots \mu(dx_{m+1}) \delta_x(dx_m) \\ \leq &2^{n-m} \|\delta_x\| \|\delta_{x'}\| \|\mu\|^{n-m-1} \leq \tilde{L}_{n-m}. \end{aligned} \tag{51}$$

Then, (34), (49) imply

$$\left| g_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi) \right| = \left| \frac{r_{\eta,\boldsymbol{y}}^{m:n}(x'|x)}{\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle} \right| \leq 2\tilde{L}_{n-m}^2 \leq \tilde{M}_{n-m} \tag{52}$$

(as $\left| \left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle \right| \geq \mathrm{Re}\left(\left\langle R_{\eta,\boldsymbol{y}}^{m:n}(\xi) \right\rangle \right) > 0$). Consequently, (35) yields

$$\begin{aligned} \left| f_{\eta,\boldsymbol{y}}^{m:n}(x|\xi) \right| \leq &\int \left| g_{\eta,\boldsymbol{y}}^{m:n}(x|x',\xi) \right| |\xi|(dx') \\ \leq &\tilde{M}_{n-m} \|\xi\| \leq 2\tilde{M}_{n-m} \leq M_{n-m} \end{aligned} \tag{53}$$

(notice that $\|\xi\| \leq 1 + \alpha_{n-m} \leq 2$ due to $\xi \in V_{\alpha_{n-m}}(\mathcal{P}(\mathcal{X}))$). Similarly, we have

$$\int \left| g_{\eta,\boldsymbol{y}}^{m:n}(x'|x,\xi) \right| \mu(dx') \leq \tilde{M}_{n-m} \|\mu\|. \tag{54}$$

Combining this with (36), (52), (53), we get

$$\begin{aligned}\left|h^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi)\right| \leq& \left|f^{m:n}_{\eta,\boldsymbol{y}}(x'|\xi)\right| \int \left|g^{m:n}_{\eta,\boldsymbol{y}}(x''|x,\xi)\right| \mu(dx'') \\ &+ \left|g^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi)\right| \\ \leq& \tilde{M}_{n-m} + 2\tilde{M}^2_{n-m}\|\mu\| \leq M_{n-m}. \end{aligned} \tag{55}$$

Due to Lemma A1.1 and (52), we have

$$\begin{aligned}\left|g^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi) - g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi)\right| \leq& \frac{2d\tilde{L}^2_{n-m}\|\eta'-\eta''\|}{\alpha_{n-m}} \\ \leq& \tilde{M}_{n-m}\|\eta'-\eta''\| \end{aligned} \tag{56}$$

(notice that $g^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi)$ is analytic in η for each $\eta \in V_{\alpha_{n-m}}(\Theta)$). Moreover, (46), (49), (52) yield

$$\begin{aligned}&\left|g^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi') - g^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi'')\right| \\ &= \left|g^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi')\right| \left|\frac{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi')\right\rangle - \left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\rangle}{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\rangle}\right| \\ &\leq 2\tilde{L}^4_{n-m}\|\xi'-\xi''\| \leq \tilde{M}_{n-m}\|\xi'-\xi''\|. \end{aligned} \tag{57}$$

Combining (56), (57), we get

$$\begin{aligned}&\left|g^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right| \\ &\leq \left|g^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi')\right| \\ &\quad + \left|g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right| \\ &\leq \tilde{M}_{n-m}(\|\eta'-\eta''\| + \|\xi'-\xi''\|). \end{aligned} \tag{58}$$

Consequently, (35), (52) imply

$$\begin{aligned}&\left|f^{m:n}_{\eta',\boldsymbol{y}}(x|\xi') - f^{m:n}_{\eta'',\boldsymbol{y}}(x|\xi'')\right| \\ &\leq \int \left|g^{m:n}_{\eta',\boldsymbol{y}}(x|x',\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x|x',\xi'')\right| |\xi'|(dx') \\ &\quad + \int \left|g^{m:n}_{\eta'',\boldsymbol{y}}(x|x',\xi'')\right| |\xi'-\xi''|(dx') \\ &\leq \tilde{M}_{n-m}\|\xi'\|(\|\eta'-\eta''\| + \|\xi'-\xi''\|) + \tilde{M}_{n-m}\|\xi'-\xi''\| \\ &\leq 3\tilde{M}_{n-m}(\|\eta'-\eta''\| + \|\xi'-\xi''\|) \\ &\leq M_{n-m}(\|\eta'-\eta''\| + \|\xi'-\xi''\|) \end{aligned} \tag{59}$$

(notice that $\|\xi'\| \leq 1 + \alpha_{n-m} \leq 2$). Similarly, we get

$$\begin{aligned}&\int \left|g^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right| \mu(dx') \\ &\leq \tilde{M}_{n-m}\|\mu\|(\|\eta'-\eta''\| + \|\xi'-\xi''\|). \end{aligned}$$

Combining this with (36), (53), (54), (58), (59), we get

$$\begin{aligned}&\left|h^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi') - h^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right| \\ &\leq \left|g^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right| \\ &\quad + \left|f^{m:n}_{\eta',\boldsymbol{y}}(x'|\xi') - f^{m:n}_{\eta'',\boldsymbol{y}}(x'|\xi'')\right| \int \left|g^{m:n}_{\eta',\boldsymbol{y}}(x''|x,\xi')\right| \mu(dx'') \\ &\quad + \left|f^{m:n}_{\eta'',\boldsymbol{y}}(x'|\xi'')\right| \int \left|g^{m:n}_{\eta',\boldsymbol{y}}(x''|x,\xi') - g^{m:n}_{\eta'',\boldsymbol{y}}(x''|x,\xi'')\right| \mu(dx'') \\ &\leq (\tilde{M}_{n-m} + 5\tilde{M}^2_{n-m}\|\mu\|)(\|\eta'-\eta''\| + \|\xi'-\xi''\|) \\ &\leq M_{n-m}(\|\eta'-\eta''\| + \|\xi'-\xi''\|). \end{aligned} \tag{60}$$

Using (49), (53), (55) – (60), we conclude that (iii), (iv) hold. □

Lemma 6.4. *Let Assumptions 2.2 and 2.3 hold. Then, the following is true:*

(i) There exists a real number $\delta_4 \in (0, \delta_1]$ such that $\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi)\} > 0$ *for all $\eta \in V_{\delta_4}(\Theta)$, $\xi \in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$, $y \in \mathcal{Y}$ (δ_1 is specified in Lemma 5.2).*

(ii) There exists a real number $C_4 \in [1, \infty)$ such that

$$\begin{aligned}&|\Phi_{\eta,y}(\xi)| \leq C_4(1+\psi(y)), \\ &|\Phi_{\eta',y}(\xi') - \Phi_{\eta'',y}(\xi'')| \\ &\leq C_4(1+\psi(y))(\|\eta'-\eta''\| + \|\xi'-\xi''\|) \end{aligned}$$

for all $\eta, \eta', \eta'' \in V_{\delta_4}(\Theta)$, $\xi, \xi', \xi'' \in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$, $y \in \mathcal{Y}$.

Proof. Throughout the proof, the following notations is used. δ_4, C_4 are the real numbers defined by $\delta_4 = \alpha_1$, $C_4 = 4L_1^2$ (α_1, L_1 are specified in Lemma 6.3). η, η', η'' are any elements of $V_{\delta_4}(\Theta)$, while ξ, ξ', ξ'' are any elements in $V_{\delta_4}(\mathcal{P}(\mathcal{X}))$. y is any element of $\mathcal{Y}$.

Using Lemma 6.3, we conclude

$$\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi)\} \geq \frac{|\varphi_\eta(y)|}{L_1}, \tag{61}$$

$$|R_{\eta,y}(\mathcal{X}|\xi)| \leq L_1|\varphi_\eta(y)|, \tag{62}$$

$$|R_{\eta,y}(\mathcal{X}|\xi') - R_{\eta,y}(\mathcal{X}|\xi'')| \leq L_1|\varphi_\eta(y)|\|\xi'-\xi''\| \tag{63}$$

(notice that $R_{\eta,y}(\mathcal{X}|y) = \left\langle R^{0:1}_{\eta,\boldsymbol{y}}(\xi)\right\rangle$ when $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ is a sequence in $\mathcal{Y}$ satisfying $y = y_1$). We also deduce that $R_{\eta,y}(\mathcal{X}|\xi)$ is analytic in η for each $\eta \in V_{\delta_4}(\Theta)$. As $\varphi_\eta(y) \neq 0$ (owing to Assumption 2.2), (61) implies that (i) holds. Consequently, (29) yields that $\Phi_{\eta,y}(\xi)$ is analytic in η for all $\eta \in V_{\delta_4}(\Theta)$. Moreover, due to (61), (62), we have

$$\begin{aligned}&\log|R_{\eta,y}(\mathcal{X}|\xi)| \leq \log L_1 + \log|\varphi_\eta(y)| \leq L_1(1+\psi(y)), \\ &\log|R_{\eta,y}(\mathcal{X}|\xi)| \geq -\log L_1 + \log|\varphi_\eta(y)| \geq -L_1(1+\psi(y)). \end{aligned}$$

Therefore, we get

$$\begin{aligned}|\Phi_{\eta,y}(\xi)| = |\log R_{\eta,y}(\mathcal{X}|\xi)| \leq& |\log|R_{\eta,y}(\mathcal{X}|\xi)|| + \pi \\ \leq& 4L_1(1+\psi(y)) \\ \leq& C_4(1+\psi(y)). \end{aligned} \tag{64}$$

Then, Lemma A1.1 implies

$$\begin{aligned}|\Phi_{\eta',y}(\xi) - \Phi_{\eta'',y}(\xi)| \leq& \frac{4dL_1(1+\psi(y))\|\eta'-\eta''\|}{\delta_4} \\ \leq& C_4(1+\psi(y))\|\eta'-\eta''\|. \end{aligned} \tag{65}$$

Let $\phi_{\eta,y}(t|\xi',\xi'')$ be the function defined by

$$\phi_{\eta,y}(t|\xi',\xi'') = \log\left(tR_{\eta,y}(\mathcal{X}|\xi') + (1-t)R_{\eta,y}(\mathcal{X}|\xi'')\right)$$

for $t \in [0,1]$. Due to Assumption 2.2 and (61), we have

$$\begin{aligned}&|tR_{\eta,y}(\mathcal{X}|\xi') + (1-t)R_{\eta,y}(\mathcal{X}|\xi'')| \\ &\geq t\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi')\} + (1-t)\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi'')\} \\ &\geq \frac{|\varphi_\eta(y)|}{L_1} > 0 \end{aligned} \tag{66}$$

for $t \in [0,1]$. Hence, $\phi_{\eta,y}(t|\xi',\xi'')$ is well-defined and differentiable in t for each $t \in [0,1]$. We also have

$$\begin{aligned}\phi'_{\eta,y}(t|\xi',\xi'') =& \frac{\partial}{\partial t}\phi_{\eta,y}(t|\xi',\xi'') \\ =& \frac{\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi')\} - \text{Re}\{R_{\eta,y}(\mathcal{X}|\xi'')\}}{t\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi')\} + (1-t)\text{Re}\{R_{\eta,y}(\mathcal{X}|\xi'')\}} \end{aligned}$$

Consequently, (63), (66) yield

$$\left|\phi'_{\eta,y}(t|\xi',\xi'')\right| \leq L_1^2\|\xi'-\xi''\|.$$

Thus, we get

$$\begin{aligned}|\Phi_{\eta,y}(\xi')-\Phi_{\eta,y}(\xi'')| &= |\phi_{\eta,y}(1|\xi',\xi'')-\phi_{\eta,y}(0|\xi',\xi'')| \\ &= \left|\int_0^1 \phi'_{\eta,y}(t|\xi',\xi'')dt\right| \\ &\leq L_1^2\|\xi'-\xi''\| \leq C_4\|\xi'-\xi''\|.\end{aligned}$$

Consequently, (65) implies

$$\begin{aligned}|\Phi_{\eta',y}(\xi')-\Phi_{\eta'',y}(\xi'')| \leq& |\Phi_{\eta',y}(\xi')-\Phi_{\eta'',y}(\xi')| \\ &+ |\Phi_{\eta'',y}(\xi')-\Phi_{\eta'',y}(\xi'')| \\ \leq& C_4(1+\psi(y))(\|\eta'-\eta''\|+\|\xi'-\xi''\|)\,.\end{aligned} \tag{67}$$

Using (64), (67), we deduce that (ii) is true. □

Lemma 6.5. *Let Assumptions 2.1 – 2.3 hold. Then, the following is true:*

(i) There exist real numbers $\delta_5, \delta_6 \in (0,\delta_4]$, $C_5 \in [1,\infty)$ and an integer $n_0 \geq 1$ such that $\text{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\right\} > 0$*, $F^{m:n}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$, $F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ and*

$$\left\|F^{m:n}_{\eta,\boldsymbol{y}}(\xi')-F^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\| \leq C_5\|\xi'-\xi''\|, \tag{68}$$

$$\left\|F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi')-F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi'')\right\| \leq \frac{\|\xi'-\xi''\|}{2} \tag{69}$$

for all $\eta \in V_{\delta_5}(\Theta)$, $\xi,\xi',\xi'' \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $m+n_0 \geq n \geq m \geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$ (δ_4 is specified in Lemma 6.4).

(ii) There exist real numbers $\delta_7 \in (0,\delta_5]$, $\delta_8 \in (0,\delta_6]$ such that $F^{m:n}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ for all $\eta \in V_{\delta_7}(\Theta)$, $\xi \in V_{\delta_8}(\mathcal{P}(\mathcal{X}))$, $m+n_0 \geq n \geq m \geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ in $\mathcal{Y}$.

Proof. (i) Throughout this part of the proof, the following notations is used. n_0 is the integer defined by

$$n_0 = \left\lceil \frac{\log(4C_3)}{|\log\gamma_2|} \right\rceil,$$

while C_5, δ_5, δ_6 are the real numbers defined by

$$C_5 = M_{n_0}(1+\|\mu\|), \quad \delta_5 = \frac{\min\{\alpha_{n_0},\delta_4\}}{16C_5^2}, \quad \delta_6 = 2C_5\delta_5 \tag{70}$$

(γ_2, C_3, α_n, M_n are specified in Lemmas 6.2, 6.3). η, η', η'' are any elements in $V_{\delta_5}(\Theta)$, while θ is any element of Θ satisfying $\|\eta-\theta\| < \delta_5$. ξ, ξ', ξ'' are any elements in $V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, while λ is any element of $\mathcal{P}(\mathcal{X})$ satisfying $\|\xi-\lambda\| < \delta_6$. x is any element of $\mathcal{X}$, while $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ is any sequence in $\mathcal{Y}$. B is any element of $\mathcal{B}(\mathcal{X})$. m, n are any integers satisfying $m+n_0 \geq n > m > 0$.

Owing to Lemma 6.3, we have $\text{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\right\} > 0$ (notice that $\delta_5 \leq \delta_6 \leq \alpha_{n_0} \leq \alpha_{n-m}$ as $n-m \leq n_0$). Hence, we get

$$\begin{aligned}&\text{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(t\xi'+(1-t)\xi'')\right\rangle\right\} \\ &= t\text{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi')\right\rangle\right\} + (1-t)\text{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\rangle\right\} > 0\end{aligned} \tag{71}$$

for $t \in [0,1]$. On the other side, using Lemma 6.3, we conclude

$$\begin{aligned}&\left|\int I_B(x')\left(h^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi')-h^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right)\mu(dx')\right| \\ &\leq \int I_B(x')\left|h^{m:n}_{\eta',\boldsymbol{y}}(x'|x,\xi')-h^{m:n}_{\eta'',\boldsymbol{y}}(x'|x,\xi'')\right|\mu(dx') \\ &\leq M_{n-m}\|\mu\|\left(\|\eta'-\eta''\|+\|\xi'-\xi''\|\right) \\ &\leq C_5\left(\|\eta'-\eta''\|+\|\xi'-\xi''\|\right)\end{aligned} \tag{72}$$

(notice that $C_5 \geq M_{n_0} \geq M_{n-m}$ as $n-m \leq n_0 \geq$). Relying on the same lemma, we deduce

$$\begin{aligned}&\left|F^{m:n}_{\eta',\boldsymbol{y}}(B|\xi')-F^{m:n}_{\eta'',\boldsymbol{y}}(B|\xi'')\right| \\ &\leq \int I_B(x)\left|f^{m:n}_{\eta',\boldsymbol{y}}(x|\xi')-f^{m:n}_{\eta'',\boldsymbol{y}}(x|\xi'')\right|\mu(dx) \\ &\leq M_{n-m}\|\mu\|\left(\|\eta'-\eta''\|+\|\xi'-\xi''\|\right) \\ &\leq C_5\left(\|\eta'-\eta''\|+\|\xi'-\xi''\|\right)\end{aligned}$$

Hence, we have

$$\left\|F^{m:n}_{\eta',\boldsymbol{y}}(\xi')-F^{m:n}_{\eta'',\boldsymbol{y}}(\xi'')\right\| \leq C_5\left(\|\eta'-\eta''\|+\|\xi'-\xi''\|\right) \tag{73}$$

(notice that $\|\eta-\theta\| < \delta_5$, $\|\xi-\lambda\| < \delta_6$). Consequently, (68) holds (set $\eta'=\eta$, $\eta''=\eta$ in (73)). We also get

$$\begin{aligned}\left\|F^{m:n}_{\eta,\boldsymbol{y}}(\xi)-F^{m:n}_{\theta,\boldsymbol{y}}(\lambda)\right\| \leq& C_5\left(\|\eta-\theta\|+\|\xi-\lambda\|\right) \\ <& C_5(\delta_5+\delta_6) \leq \delta_4\end{aligned}$$

Therefore, $F^{m:n}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$ for $m+n_0 \geq n > m \geq 0$ (notice that $F^{m:n}_{\theta,\boldsymbol{y}}(\lambda) \in \mathcal{P}(\mathcal{X})$).

Let $\phi^{m:n}_{\eta,\boldsymbol{y}}(t,x|\xi',\xi'')$ be the function defined by

$$\phi^{m:n}_{\eta,\boldsymbol{y}}(t,x|\xi',\xi'') = f^{m:n}_{\eta,\boldsymbol{y}}(x|t\xi'+(1-t)\xi'') \tag{74}$$

for $t \in [0,1]$, $m+n_0 \geq n > m \geq 0$. Then, due to (71), we have

$$\begin{aligned}&\phi^{m:n}_{\eta,\boldsymbol{y}}(t,x|\xi',\xi'') \\ &= \frac{\int r^{m:n}_{\eta,\boldsymbol{y}}(x|x')(t\xi'+(1-t)\xi'')(dx')}{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(t\xi'+(1-t)\xi'')\right\rangle} \\ &= \frac{t\int r^{m:n}_{\eta,\boldsymbol{y}}(x|x')\xi'(dx')+(1-t)\int r^{m:n}_{\eta,\boldsymbol{y}}(x|x')\xi''(dx')}{t\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi')\right\rangle+(1-t)\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\rangle}\end{aligned}$$

Thus, we get

$$\begin{aligned}\frac{\partial}{\partial t}\phi^{m:n}_{\eta,\boldsymbol{y}}(t,x|\xi',\xi'') =& \frac{\int r^{m:n}_{\eta,\boldsymbol{y}}(x|x')(\xi'-\xi'')(dx')}{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(t\xi'+(1-t)\xi'')\right\rangle} \\ &- f^{m:n}_{\eta,\boldsymbol{y}}(x|t\xi'+(1-t)\xi'') \\ &\cdot \frac{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi')\right\rangle-\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\rangle}{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(t\xi'+(1-t)\xi'')\right\rangle} \\ =& \frac{\int r^{m:n}_{\eta,\boldsymbol{y}}(x|x')(\xi'-\xi'')(dx')}{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(t\xi'+(1-t)\xi'')\right\rangle} \\ &- f^{m:n}_{\eta,\boldsymbol{y}}(x|t\xi'+(1-t)\xi'') \\ &\cdot \frac{\int\int r^{m:n}_{\eta,\boldsymbol{y}}(x''|x')\mu(dx'')(\xi'-\xi'')(dx')}{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(t\xi'+(1-t)\xi'')\right\rangle}\end{aligned}$$

Consequently, (34) – (36) imply

$$\begin{aligned}&\frac{\partial}{\partial t}\phi^{m:n}_{\eta,\boldsymbol{y}}(t,x|\xi',\xi'')\\&=\int g^{m:n}_{\eta,\boldsymbol{y}}(x|x',t\xi'+(1-t)\xi'')(\xi'-\xi'')(dx')\\&\quad-f^{m:n}_{\eta,\boldsymbol{y}}(x|t\xi'+(1-t)\xi'')\\&\quad\cdot\int\int g^{m:n}_{\eta,\boldsymbol{y}}(x''|x',t\xi'+(1-t)\xi'')\mu(dx'')(\xi'-\xi'')(dx')\\&=\int h^{m:n}_{\eta,\boldsymbol{y}}(x|x',t\xi'+(1-t)\xi'')(\xi'-\xi'')(dx')\end{aligned}\tag{75}$$

Hence, we have

$$\begin{aligned}&f^{m:n}_{\eta,\boldsymbol{y}}(x|\xi')-f^{m:n}_{\eta,\boldsymbol{y}}(x|\xi'')\\&=\phi^{m:n}_{\eta,\boldsymbol{y}}(1,x|\xi',\xi'')-\phi^{m:n}_{\eta,\boldsymbol{y}}(0,x|\xi',\xi'')\\&=\int\int_0^1 h^{m:n}_{\eta,\boldsymbol{y}}(x|x',t\xi'+(1-t)\xi'')(\xi'-\xi'')(dx')dt\end{aligned}\tag{76}$$

Therefore, (37) yields

$$\begin{aligned}&F^{m:n}_{\eta,\boldsymbol{y}}(B|\xi')-F^{m:n}_{\eta,\boldsymbol{y}}(B|\xi'')\\&=\int I_B(x)\left(f^{m:n}_{\eta,\boldsymbol{y}}(x|\xi')-f^{m:n}_{\eta,\boldsymbol{y}}(x|\xi'')\right)\mu(dx)\\&=\int\int\int_0^1 I_B(x)h^{m:n}_{\eta,\boldsymbol{y}}(x|x',t\xi'+(1-t)\xi'')\\&\quad\cdot\mu(dx)(\xi'-\xi'')(dx')dt\end{aligned}\tag{77}$$

Consequently, we have

$$\begin{aligned}&F^{m:n}_{\theta,\boldsymbol{y}}(B|\lambda+\alpha\delta_x)-F^{m:n}_{\theta,\boldsymbol{y}}(B|\lambda)\\&=\alpha\int\int_0^1 I_B(x')h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda+\alpha t\delta_x)\mu(dx')dt\end{aligned}\tag{78}$$

for $\alpha\in(0,\delta_6)$ (notice $\lambda+\alpha t\delta_x\in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ when $\alpha\in(0,\delta_6)$, $t\in[0,1]$). Moreover, Lemma 6.2 yields

$$\begin{aligned}&\left|F^{m:n}_{\theta,\boldsymbol{y}}(B|\lambda+\alpha\delta_x)-F^{m:n}_{\theta,\boldsymbol{y}}(B|\lambda)\right|\\&\leq C_3\gamma_2^{n-m}\|\alpha\delta_x\|=\alpha C_3\gamma_2^{n-m}\end{aligned}\tag{79}$$

Combining (78), (79), we thus get

$$\left|\int\int_0^1 I_B(x')h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda+\alpha t\delta_x)\mu(dx')dt\right|\leq C_3\gamma_2^{n-m}.\tag{80}$$

Using (72), (80), we conclude

$$\begin{aligned}&\left|\int I_B(x')h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda)\mu(dx')\right|\\&\leq\left|\int\int_0^1 I_B(x')h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda+\alpha t\delta_x)\mu(dx')dt\right|\\&\quad+\int_0^1\left|\int I_B(x')\left(h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda+\alpha t\delta_x)-h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda)\right)\mu(dx')\right|dt\\&\leq C_3\gamma_2^{n-m}+C_5\alpha\end{aligned}$$

for $\alpha\in(0,\delta_6)$ (as $\|\alpha t\delta_x\|\leq\alpha$). Letting $\alpha\to 0$, we deduce

$$\left|\int I_B(x')h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda)\mu(dx')\right|\leq C_3\gamma_2^{n-m}.$$

Consequently, (72) yields

$$\begin{aligned}&\left|\int I_B(x')h^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi)\mu(dx')\right|\\&\leq\left|\int I_B(x')h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda)\mu(dx')\right|\\&\quad+\left|\int I_B(x')\left(h^{m:n}_{\eta,\boldsymbol{y}}(x'|x,\xi)-h^{m:n}_{\theta,\boldsymbol{y}}(x'|x,\lambda)\right)\mu(dx')\right|\\&\leq C_3\gamma_2^{n-m}+C_5\left(\|\eta-\theta\|+\|\xi-\lambda\|\right)\\&\leq C_3\gamma_2^{n-m}+C_5(\delta_5+\delta_6)\leq C_3\gamma_2^{n-m}+\frac{1}{4}\end{aligned}$$

(notice that $\|\eta-\theta\|<\delta_5$, $\|\xi-\lambda\|<\delta_6$, $C_5\delta_5\leq C_5\delta_6\leq 1/8$). Combining this with (77), we get

$$\begin{aligned}&\left|F^{m:n}_{\eta,\boldsymbol{y}}(B|\xi')-F^{m:n}_{\eta,\boldsymbol{y}}(B|\xi'')\right|\\&\leq\int\int_0^1\left|\int I_B(x')h^{m:n}_{\eta,\boldsymbol{y}}(x'|x,t\xi'+(1-t)\xi'')\mu(dx')\right|\\&\quad\cdot|\xi'-\xi''|(dx)dt\\&\leq\left(C_3\gamma_2^{n-m}+\frac{1}{4}\right)\|\xi'-\xi''\|\end{aligned}$$

(notice that $t\xi'+(1-t)\xi''\in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ as $\xi',\xi''\in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ and $V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ is convex). Therefore, we have

$$\left\|F^{m:n}_{\eta,\boldsymbol{y}}(\xi')-F^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\|\leq\left(C_3\gamma_2^{n-m}+\frac{1}{4}\right)\|\xi'-\xi''\|.$$

Hence, we get

$$\begin{aligned}\left\|F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi')-F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi'')\right\|&\leq\left(C_3\gamma_2^{n_0}+\frac{1}{4}\right)\|\xi'-\xi''\|\\&\leq\frac{\|\xi'-\xi''\|}{2}\end{aligned}\tag{81}$$

(notice that $C_3\gamma_2^{n_0}\leq 1/4$). Consequently, (69) holds. Moreover, (73) implies

$$\begin{aligned}&\left\|F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi)-F^{m:m+n_0}_{\theta,\boldsymbol{y}}(\lambda)\right\|\\&\leq\left\|F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi)-F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\lambda)\right\|\\&\quad+\left\|F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\lambda)-F^{m:m+n_0}_{\theta,\boldsymbol{y}}(\lambda)\right\|\\&\leq\frac{\|\xi-\lambda\|}{2}+C_5\|\eta-\theta\|\\&<\frac{\delta_6}{2}+C_5\delta_5=\delta_6\end{aligned}$$

(notice that $\|\eta-\theta\|<\delta_5$, $\|\xi-\lambda\|<\delta_6$,). Thus, $F^{m:m+n_0}_{\eta,\boldsymbol{y}}(\xi)\in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ for $m\geq 0$ (notice that $F^{m:m+n_0}_{\theta,\boldsymbol{y}}(\lambda)\in\mathcal{P}(\mathcal{X})$).

(ii) Let δ_7, δ_8 be the real numbers defined by $\delta_7=\delta_5$, $\delta_8=\delta_5$ (δ_5 is specified in (70)). Moreover, let θ, λ, $\boldsymbol{y}$ have the same meaning as in (i), while η, ξ are any elements of $V_{\delta_6}(\Theta)$, $V_{\delta_7}(\mathcal{P}(\mathcal{X}))$ (respectively). Consequently, when $\|\eta-\theta\|<\delta_7$, $\|\xi-\lambda\|<\delta_8$, (73) yields

$$\begin{aligned}\left\|F^{m:n}_{\eta,\boldsymbol{y}}(\xi)-F^{m:n}_{\theta,\boldsymbol{y}}(\lambda)\right\|&\leq C_5\left(\|\eta-\theta\|+\|\xi-\lambda\|\right)\\&<C_5(\delta_7+\delta_8)\leq\delta_6\end{aligned}$$

for $m+n_0\geq n>m\geq 0$. Therefore, $F^{m:n}_{\eta,\boldsymbol{y}}(\xi)\in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ for $m+n_0\geq n>m\geq 0$. □

Lemma 6.6. *Let Assumptions 2.1 – 2.3 hold. Then, the following is true:*

(i) Re $\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\right\} \neq 0$, $F^{m:n}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$ *for all* $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n \geq m \geq 0$ *and any sequence* $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ *in* $\mathcal{Y}$ *(*δ_4, δ_5, δ_6 *are specified in Lemmas 6.4, 6.5).*

(ii) There exist real numbers $\gamma_3 \in (0,1)$, $C_6 \in [1,\infty)$ *such that*

$$\left\|F^{m:n}_{\eta,\boldsymbol{y}}(\xi') - F^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\| \leq C_6\gamma_3^{n-m}\left\|\xi' - \xi''\right\| \tag{82}$$

for all $\eta \in V_{\delta_5}(\Theta)$, $\xi', \xi'' \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n \geq m \geq 0$ *and any sequence* $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ *in* $\mathcal{Y}$.

Proof. (i) Let $n_k(m)$ be the integer defined by $n_k(m) = m + kn_0$ for $m, k \geq 0$ (n_0 is specified in Lemma 6.5). Moreover, let $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ be any sequence in $\mathcal{Y}$.

First, we show

$$\text{Re}\left\{\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\right\} \neq 0, \tag{83}$$

$$F^{m:n}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_4}(\mathcal{P}(\mathcal{X})), \tag{84}$$

$$F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_6}(\mathcal{P}(\mathcal{X})) \tag{85}$$

for each $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_k(m) \geq n \geq m \geq 0$, $k \geq 0$. We prove this by induction in k.

We have $R^{m:n}_{\eta,\boldsymbol{y}}(\xi) = \xi$, $F^{m:n}_{\eta,\boldsymbol{y}}(\xi) = \xi$ for $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_0(m) \geq n \geq m \geq 0$ (notice that $n_0(m) = n = m$ when $n_0(m) \geq n \geq m \geq 0$). Hence, (83) – (85) hold for $k = 0$ and $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_k(m) \geq n \geq m \geq 0$. Now, the induction hypothesis is formulated: Suppose that (83) – (85) are true for some $k \geq 0$ and any $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_k(m) \geq n \geq m \geq 0$. Then, to show (83) – (85) for $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_{k+1}(m) \geq n \geq m \geq 0$, it is sufficient to demonstrate (83) – (85) for $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_{k+1}(m) \geq n \geq n_k(m)$, $m \geq 0$.

In the rest of the proof of (i), η, ξ are any elements of $V_{\delta_5}(\Theta)$, $V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ (respectively). Owing to Lemma 6.5, we have Re $\left\{\left\langle R^{n_k(m):n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\right\} > 0$, $F^{n_k(m):n}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$, $F^{n_k(m):n_{k+1}(m)}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ for $n_{k+1}(m) \geq n \geq n_k(m)$, $m \geq 0$. Combining this with the induction hypothesis, we get

$$\text{Re}\left\{\left\langle R^{n_k(m):n}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right)\right\rangle\right\} > 0,$$

$$F^{n_k(m):n}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right) \in V_{\delta_4}(\mathcal{P}(\mathcal{X})), \tag{86}$$

$$F^{n_k(m):n_{k+1}(m)}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right) \in V_{\delta_6}(\mathcal{P}(\mathcal{X})) \tag{87}$$

for $n_{k+1}(m) \geq n \geq n_k(m)$, $m \geq 0$ (notice that $F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ follows from the induction hypothesis). Then, the induction hypothesis and Lemma 6.1 (Part (ii)) imply

$$\begin{aligned}\left\langle R^{m:n}_{\eta,\boldsymbol{y}}(\xi)\right\rangle =& \left\langle R^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right\rangle\left\langle R^{n_k(m):n}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right)\right\rangle \\ \neq& 0\end{aligned} \tag{88}$$

for $n_{k+1}(m) \geq n \geq n_k(m)$, $m \geq 0$ (notice that $\left\langle R^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right\rangle \neq 0$ results from the induction hypothesis). Therefore, Lemma 6.1 (Part (iii)) and (86), (87) yield

$$F^{m:n}_{\eta,\boldsymbol{y}}(\xi) = F^{n_k(m):n}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right) \in V_{\delta_4}(\mathcal{P}(\mathcal{X})), \tag{89}$$

$$\begin{aligned}F^{m:n_{k+1}(m)}_{\eta,\boldsymbol{y}}(\xi) =& F^{n_k(m):n_{k+1}(m)}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right) \\ \in& V_{\delta_6}(\mathcal{P}(\mathcal{X}))\end{aligned} \tag{90}$$

for $n_{k+1}(m) \geq n \geq n_k(m)$, $m \geq 0$ (notice that $\left\langle R^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi)\right\rangle \neq 0$, $\left\langle R^{m:n_{k+1}(m)}_{\eta,\boldsymbol{y}}(\xi)\right\rangle \neq 0$, $F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi) \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ are a consequence of the induction hypothesis and (88)). Combining (88) – (90) with the induction hypothesis, we deduce that (83) – (85) hold for $n_{k+1}(m) \geq n \geq m$, $m \geq 0$. Then, relying on the principle of mathematical induction, we conclude that (83) – (85) are satisfied for each $\eta \in V_{\delta_5}(\Theta)$, $\xi \in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$, $n_k(m) \geq n \geq m \geq 0$, $k \geq 0$. As a direct consequence of this, we have that (i) is true.

(ii) Let γ_3, C_6 be the real numbers defined by $\gamma_3 = 2^{-1/n_0}$, $C_6 = C_5\gamma_3^{-n_0}$ (C_5, n_0 are specified in Lemma 6.5), while $n_k(m)$, $\boldsymbol{y}$ have the same meaning as in (i). Moreover, let η be any element of $V_{\delta_6}(\Theta)$, while ξ', ξ'' are any elements in $V_{\delta_6}(\mathcal{P}(\mathcal{X}))$.

Owing to Lemmas 6.1, 6.5 and (83) – (85), we have

$$\begin{aligned}&\left\|F^{m:n_{k+1}(m)}_{\eta,\boldsymbol{y}}(\xi') - F^{m:n_{k+1}(m)}_{\eta,\boldsymbol{y}}(\xi'')\right\| \\ &= \Big\|F^{n_k(m):n_{k+1}(m)}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi')\right) \\ &\quad - F^{n_k(m):n_{k+1}(m)}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi'')\right)\Big\| \\ &\leq \frac{1}{2}\left\|F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi') - F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi'')\right\|\end{aligned}$$

for $m, k \geq 0$. Consequently, we have

$$\begin{aligned}&\left\|F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi') - F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi'')\right\| \\ &\leq \frac{1}{2^k}\left\|F^{m:m}_{\eta,\boldsymbol{y}}(\xi') - F^{m:m}_{\eta,\boldsymbol{y}}(\xi'')\right\| = \gamma_3^{n_k(m)-m}\|\xi' - \xi''\|\end{aligned}$$

Combining this with Lemma 6.5 and (83) – (85), we get

$$\begin{aligned}&\left\|F^{m:n}_{\eta,\boldsymbol{y}}(\xi') - F^{m:n}_{\eta,\boldsymbol{y}}(\xi'')\right\| \\ &= \left\|F^{n_k(m):n}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi')\right) - F^{n_k(m):n}_{\eta,\boldsymbol{y}}\left(F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi'')\right)\right\| \\ &\leq C_5\left\|F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi') - F^{m:n_k(m)}_{\eta,\boldsymbol{y}}(\xi'')\right\| \\ &\leq C_5\gamma_3^{n_k(m)-m}\|\xi' - \xi''\| \leq C_6\gamma_3^{n-m}\|\xi' - \xi''\|\end{aligned} \tag{91}$$

for $n_{k+1}(m) > n \geq n_k(m)$, $m, k \geq 0$ (notice that $C_5\gamma_3^{n_k(m)-m} = (C_5\gamma_3^{n_k(m)-n})\gamma_3^{n-m} \leq (C_5\gamma_3^{-n_0})\gamma_3^{n-m} = C_6\gamma_3^{n-m}$). Thus, (82) holds for each $n \geq m \geq 0$ (set $k = \lfloor (n-m)/n_0 \rfloor$ in (91)). ☐

VII. Proof of Main Results

In this section, Theorems 2.1 and 2.2 are proved. The proofs of these theorems crucially depend on the results related to the kernels $S(z, dz')$, $S_\eta(z, dz')$ and the optimal filter $F^{m:n}_{\eta,\boldsymbol{y}}(\xi)$ (i.e., on Lemmas 5.1, 5.4, 6.6). As the properties of $S(z, dz')$, $S_\eta(z, dz')$ are very similar, the proofs of Theorems 2.1 and 2.2 have many elements in common. In order not to consider these elements twice (and to prove Theorems 2.1 and 2.2 as efficiently as possible), we introduce a new kernel $T_\eta(z,dz')$, where $\eta\in\mathbb{C}^d$, $z\in\mathcal{Z}$.[1] Its purpose is to capture all common features of $S(z,dz')$, $S_\eta(z,dz')$ which are relevant for the proof of Theorems 2.1 and 2.2. Using $T_\eta(z,dz')$, we recursively define kernels $\left\{T_\eta^n(z,dz')\right\}_{n\geq 0}$ by $T_\eta^0(z,B)=\delta_z(B)$ and

$$T_\eta^{n+1}(z,B)=\int T_\eta^n(z',B)T_\eta(z,dz'),$$

where $B\in\mathcal{B}(\mathcal{Z})$.

Regarding $T_\eta(z,dz')$, we assume the following.

Assumption 7.1. *For each $\theta\in\Theta$, $z\in\mathcal{Z}$, $T_\theta(z,dz')$ is a probability measure.*

Assumption 7.2. *(i) There exist real numbers $\alpha\in(0,\delta]$, $L\in[1,\infty)$ such that*

$$\left|T_{\eta'}-T_{\eta''}\right|(z,B)\leq L\|\eta'-\eta''\|,$$
$$\int\tilde{\psi}(z')|T_\eta|(z,dz')\leq L$$

for all $\eta,\eta',\eta''\in V_\alpha(\Theta)$, $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$ (here, $|T_{\eta'}-T_{\eta''}|(z,dz')$ denotes the total variation of $T_{\eta'}(z,dz')-T_{\eta''}(z,dz')$, while δ, $\tilde{\psi}(z)$ are specified in Assumption 2.2 and (11)).

(ii) For each $\eta\in V_\alpha(\Theta)$, there exists a complex measure $\tau_\eta(dz)$ such that $\lim_{n\to\infty}T_\eta^n(z,B)=\tau_\eta(B)$ for all $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$.

(iii) There exists a real number $\beta\in(0,1)$ such that

$$\left|T_\eta^n-\tau_\eta\right|(z,B)\leq L\beta^n$$

for all $\eta\in V_\alpha(\Theta)$, $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$, $n\geq 1$ (here, $\left|T_\eta^n-\tau_\eta\right|(z,dz')$ stands for the total variation of $T_\eta^n(z,dz')-\tau_\eta(dz')$).

Remark. *According to Lemmas 5.1 and 5.4, both kernels $S(z,dz')$, $S_\eta(z,dz')$ satisfy Assumptions 7.1 and 7.2. These assumptions capture all common properties of $S(z,dz')$, $S_\eta(z,dz')$ relevant for the proof of Theorems 2.1 and 2.2.*

Besides the notations introduced in the previous sections, we rely here on the following notations, too. $u_\eta^n(z_{0:n})$ and $F_\eta^n(\xi,z_{1:n})$ are (respectively) the function and the complex measure defined by

$$u_\eta^n(z_{0:n})=u_\eta^n(x_{0:n},y_{1:n}),\quad F_\eta^n(\xi,z_{1:n})=F_{\eta,\boldsymbol{y}}^{0:n}(\xi)\tag{92}$$

for $\eta\in\mathbb{C}^d$, $\xi\in\mathcal{M}_c(\mathcal{X})$, $x_0,\dots,x_n\in\mathcal{X}$, $y_0,\dots,y_n\in\mathcal{Y}$, $n\geq 0$ and $z_0=(y_0,x_0),\dots,z_n=(y_n,x_n)$, where $\boldsymbol{y}=\{y_n'\}_{n\geq 1}$ is any sequence in $\mathcal{Y}$ satisfying $y_k'=y_k$ for $n\geq k\geq 1$.[2] $\Phi_\eta(\xi,z)$ is the function defined by

$$\Phi_\eta(\xi,z)=\Phi_{\eta,y}(\xi),$$

where $x\in\mathcal{X}$, $y\in\mathcal{Y}$ and $z=(y,x)$ ($\Phi_{\eta,y}(\xi)$ is specified in (29)).[3] $\Phi_\eta^n(\xi,z)$ is the function defined by

$$\Phi_\eta^n(\xi,z)=\int\cdots\int\Phi_\eta\left(F_\eta^n(\xi,z_{1:n}),z_{n+1}\right)\cdot T_\eta(z_n,dz_{n+1})\cdots T_\eta(z,dz_1)\tag{93}$$

for $n\geq 0$. $\bar{A}_\eta^n(\xi)$, $A_\eta^{k,n}(\xi,z)$, $B_\eta^n(\xi,z)$ are the functions defined by

$$\begin{aligned}\bar{A}_\eta^n(\xi)=&\int\cdots\int\int\Big(\Phi_\eta\left(F_\eta^n(\xi,z_{1:n}),z_{n+1}\right)-\Phi_\eta\left(F_\eta^{n-1}(\xi,z_{2:n}),z_{n+1}\right)\Big)\\&\cdot T_\eta(z_n,dz_{n+1})\cdots T_\eta(z_0,dz_1)\tau_\eta(dz_0),\\A_\eta^{k,n}(\xi,z)=&\int\cdots\int\int\Big(\Phi_\eta\left(F_\eta^{n-k+1}(\xi,z_{k:n}),z_{n+1}\right)-\Phi_\eta\left(F_\eta^{n-k}(\xi,z_{k+1:n}),z_{n+1}\right)\Big)\\&\cdot T_\eta(z_n,dz_{n+1})\cdots T_\eta(z_k,dz_{k+1})(T_\eta^k-\tau_\eta)(z,dz_k),\\B_\eta^n(\xi,z)=&\int\Phi_\eta\left(\xi,z'\right)(T_\eta^{n+1}-\tau_\eta)(z,dz')\end{aligned}$$

for $n\geq k\geq 1$.

Under the notations introduced above, we have

$$\log q_\theta^n(y_{1:n}|\lambda)=\sum_{k=0}^{n-1}\Phi_\theta\left(F_\theta^k(\lambda,z_{1:k}),z_{k+1}\right)\tag{94}$$

for $\theta\in\Theta$, $\lambda\in\mathcal{P}(\mathcal{X})$, $x_1,\dots,x_n\in\mathcal{X}$, $y_1,\dots,y_n\in\mathcal{Y}$, $n\geq 1$ and $z_1=(y_1,x_1),\dots,z_n=(y_n,x_n)$. We also have

$$\Phi_\eta^n(\xi,z')-\Phi_\eta^n(\xi,z'')=\sum_{k=1}^n\left(A_\eta^{k,n}(\xi,z')-A_\eta^{k,n}(\xi,z'')\right)+B_\eta^n(\xi,z')-B_\eta^n(\xi,z''),\tag{95}$$

$$\Phi_\eta^{n+1}(\xi,z)-\Phi_\eta^n(\xi,z)=\sum_{k=1}^{n+1}A_\eta^{k,n+1}(\xi,z)-\sum_{k=1}^n A_\eta^{k,n}(\xi,z)+\bar{A}_\eta^{n+1}(\xi)+B_\eta^{n+1}(\xi,z)-B_\eta^n(\xi,z)\tag{96}$$

for $\eta\in V_\alpha(\Theta)$, $\xi\in\mathcal{M}_c(\mathcal{X})$, $z,z',z''\in\mathcal{Z}$, $n\geq 1$.

Lemma 7.1. *Let Assumptions 2.1 – 2.3, 7.1 and 7.2 hold. Then, there exist a function ϕ_η mapping $\eta\in\mathbb{C}^d$ to $\mathbb{C}$ and real numbers $\delta_9,\gamma_4\in(0,1)$, $C_7\in[1,\infty)$ such that*

$$\left|\Phi_\eta^n(\xi,z)-\phi_\eta\right|\leq C_7 n\gamma_4^n\tag{97}$$

for all $\eta\in V_{\delta_9}(\Theta)$, $\xi\in V_{\delta_9}(\mathcal{P}(\mathcal{X}))$, $z\in\mathcal{Z}$, $n\geq 1$.

Proof. Throughout the proof, the following notations is used. γ_4, δ_9 are the real numbers defined by $\gamma_4=\max\{\beta^{1/2},\gamma_3^{1/2}\}$, $\delta_9=\min\{\delta_7,\delta_8,(1-\gamma_4)/L\}$ (β, δ_7, δ_8, γ_3, L are specified in Assumption 7.2 and Lemmas 6.5, 6.6). η is any element in $V_{\delta_9}(\Theta)$, while θ is any element of Θ satisfying $\|\eta-\theta\|<\delta_9$. ξ, ξ', ξ'' are any elements of $V_{\delta_9}(\mathcal{P}(\mathcal{X}))$, while z, z', z'' are any elements in $\mathcal{Z}$. B is any element of $\mathcal{B}(\mathcal{Z})$. n, k are any integers satisfying $n\geq k\geq 1$.

[1] $T_\eta(z,dz')$ can be considered as a mapping with the following properties: (i) $T_\eta(z,B)$ maps $\eta\in\mathbb{C}^d$, $z\in\mathcal{Z}$, $B\in\mathcal{B}(\mathcal{Z})$ to $\mathbb{C}$, (ii) $T_\eta(z,B)$ is measurable in (η,z) for each $B\in\mathcal{B}(\mathcal{Z})$, and (iii) $T_\eta(z,B)$ is a complex measure in B for each $\eta\in\mathbb{C}^d$, $z\in\mathcal{Z}$.

[2] Here, $y_{1:0}$, $z_{1:0}$ denote empty sequences (i.e., sequences without any element). $u_\eta^n(x_{0:n},y_{1:n})$, $F_{\eta,\boldsymbol{y}}^{0:n}(\xi)$ are specified in (12),(37). Notice that $F_{\eta,\boldsymbol{y}}^{0:n}(\xi)$ depends only on $y_1',\dots,y_n'$ and is independent of other elements of $\boldsymbol{y}$.

[3] Notice that $u_\eta^n(z_{0:n})$, $F_\eta^n(\xi,z_{1:n})$, $\Phi_\eta(\xi,z)$ are just another notations for $u_\eta^n(x_{0:n},y_{1:n})$, $F_{\eta,\boldsymbol{y}}^{0:n}(\xi)$, $\Phi_{\eta,y}(\xi)$. However, notations $u_\eta^n(z_{0:n})$, $F_\eta^n(\xi,z_{1:n})$, $\Phi_\eta(\xi,z)$ is more suitable (than the original one) for measure-theoretic arguments which the analysis carried out in this section is based on.

Owing to Assumptions 7.1 and 7.2, we have

$$\begin{aligned}|T_\eta|(z,B) \leq& T_\theta(z,B)+|T_\eta-T_\theta|(z,B)\\ \leq& 1+L\|\eta-\theta\|\\ <& 1+L\delta_9\leq\frac{1}{\gamma_4}\end{aligned}\tag{98}$$

(as $L\delta_9 \leq 1-\gamma_4 \leq 1/\gamma_4 - 1$). Consequently, Assumption 7.1 yields

$$|\tau_\eta|(B)\leq|T_\eta-\tau_\eta|(z,B)+|T_\eta|(z,B)\leq L+\frac{1}{\gamma_4}.\tag{99}$$

Let $\tilde{C}_1 = 4C_4C_6$ (C_4, C_6 are specified in Lemmas 6.4, 6.6). Then, due to to Lemmas 6.1, 6.4, 6.6, we have

$$\begin{aligned}&\left|\Phi_\eta\left(F_\eta^{n-k+1}(\xi,z_{k:n}),z_{n+1}\right)-\Phi_\eta\left(F_\eta^{n-k}(\xi,z_{k+1:n}),z_{n+1}\right)\right|\\ &\leq C_4\tilde{\psi}(z_{n+1})\left\|F_\eta^{n-k+1}(\xi,z_{k:n})-F_\eta^{n-k}(\xi,z_{k+1:n})\right\|\\ &=C_4\tilde{\psi}(z_{n+1})\left\|F_\eta^{n-k}\left(F_\eta^1(\xi,z_k),z_{k+1:n}\right)-F_\eta^{n-k}(\xi,z_{k+1:n})\right\|\\ &\leq C_4C_6\gamma_3^{n-k}\tilde{\psi}(z_{n+1})\left\|F_\eta^1(\xi,z_k)-\xi\right\|\\ &\leq \tilde{C}_1\gamma_4^{2(n-k)}\tilde{\psi}(z_{n+1})\end{aligned}\tag{100}$$

for $z_1,\dots,z_{n+1}\in\mathcal{Z}$.[4] Similarly, owing to Lemmas 6.4, 6.6, we have

$$\begin{aligned}&\left|\Phi_\eta\left(F_\eta^n(\xi',z_{1:n}),z_{n+1}\right)-\Phi_\eta\left(F_\eta^n(\xi'',z_{1:n}),z_{n+1}\right)\right|\\ &\leq C_4\tilde{\psi}(z_{n+1})\left\|F_\eta^n(\xi',z_{1:n})-F_\eta^n(\xi'',z_{1:n})\right\|\\ &\leq C_4C_6\gamma_3^n\tilde{\psi}(z_{n+1})\|\xi'-\xi''\|\\ &\leq \tilde{C}_1\gamma_4^{2n}\tilde{\psi}(z_{n+1}).\end{aligned}\tag{101}$$

Let $\tilde{C}_2 = 2\tilde{C}_1L^2/\gamma_4^3$. Then, using Assumption 7.2 and (98), (101), we conclude

$$\begin{aligned}\left|\Phi_\eta^n(\xi',z)-\Phi_\eta^n(\xi'',z)\right|\leq&\tilde{C}_1\gamma_4^{2n}\int\cdots\int\tilde{\psi}(z_{n+1})\\ &\cdot|T_\eta|(z_n,dz_{n+1})\cdots|T_\eta|(z,dz_1)\\ \leq&\tilde{C}_1L\gamma_4^n\leq\tilde{C}_2\gamma_4^n.\end{aligned}\tag{102}$$

Similarly, relying on Assumption 7.2 and (98), (100), we deduce

$$\begin{aligned}\left|A_\eta^{k,n}(\xi,z)\right|\leq&\tilde{C}_1\gamma_4^{2(n-k)}\int\cdots\int\int\tilde{\psi}(z_{n+1})\\ &\cdot|T_\eta|(z_n,dz_{n+1})\cdots|T_\eta|(z_k,dz_{k+1})\\ &\cdot|T_\eta^k-\tau_\eta|(z,dz_k)\\ \leq&\tilde{C}_1L^2\beta^k\gamma_4^{n-k}\leq\tilde{C}_2\gamma_4^n.\end{aligned}\tag{103}$$

Moreover, using Assumption 7.2 and (99), (100), we get

$$\begin{aligned}\left|\bar{A}_\eta^n(\xi)\right|\leq&\tilde{C}_1\gamma_4^{2(n-1)}\int\cdots\int\int\tilde{\psi}(z_{n+1})\\ &\cdot|T_\eta|(z_n,dz_{n+1})\cdots|T_\eta|(z_0,dz_1)|\tau_\eta|(dz_0)\\ \leq&\tilde{C}_1L\left(L+\frac{1}{\gamma_4}\right)\gamma_4^{n-2}\leq\tilde{C}_2\gamma_4^n.\end{aligned}\tag{104}$$

[4]To get the first two relations in (100), use Lemmas 6.1, 6.4, and notice that inclusions $\eta\in V_{\delta_4}(\Theta)$, $F_\eta^{n-k+1}(\xi',z_{k:n})\in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$, $F_\eta^{n-k}(\xi'',z_{k+1:n})\in V_{\delta_4}(\mathcal{P}(\mathcal{X}))$ follow from Lemma 6.6 and $\eta\in V_{\delta_9}(\Theta)\subseteq V_{\delta_5}(\Theta)$, $\xi',\xi''\in V_{\delta_9}(\mathcal{P}(\mathcal{X}))\subseteq V_{\delta_6}(\mathcal{P}(\mathcal{X}))$. To get the third relation in (100), use Lemma 6.6 and notice that $F_\eta^1(\xi,z_k)\in V_{\delta_6}(\mathcal{P}(\mathcal{X}))$ follows from Lemma 6.4 and $\eta\in V_{\delta_9}(\Theta)\subseteq V_{\delta_7}(\Theta)$, $\xi\in V_{\delta_9}(\mathcal{P}(\mathcal{X}))\subseteq V_{\delta_8}(\mathcal{P}(\mathcal{X}))$. To get the last relation in (100), notice that $\|F_\eta^1(\xi,z_k)-\xi\|\leq\|F_\eta^1(\xi,z_k)\|+\|\xi\|\leq 2+\delta_6+\delta_9\leq 4$.

Let $\tilde{C}_3 = C_4L^2$, $\tilde{C}_4 = 4(\tilde{C}_2+\tilde{C}_3)$. Then, owing to Assumption 7.2 and Lemma 6.4, we have

$$\begin{aligned}\left|B_\eta^n(\xi,z)\right|\leq&C_4\int\int\tilde{\psi}(z'')|T_\eta|(z',dz'')|T_\eta^n-\tau_\eta|(z,dz')\\ \leq&C_4L^2\beta^n\leq\tilde{C}_3\gamma_4^n.\end{aligned}\tag{105}$$

Consequently, (96), (103), (104) yield

$$\begin{aligned}&\left|\Phi_\eta^{n+1}(\xi,z)-\Phi_\eta^n(\xi,z)\right|\\ &\leq\sum_{k=1}^{n+1}\left|A_\eta^{k,n+1}(\xi,z)\right|+\sum_{k=1}^{n}\left|A_\eta^{k,n}(\xi,z)\right|\\ &\quad+\left|\bar{A}_\eta^{n+1}(\xi)\right|+\left|B_\eta^{n+1}(\xi,z)\right|+\left|B_\eta^n(\xi,z)\right|\\ &\leq 2\tilde{C}_2(n+1)\gamma_4^n+2\tilde{C}_3\gamma_4^n\leq\tilde{C}_4n\gamma_4^n.\end{aligned}\tag{106}$$

Hence, we have

$$\begin{aligned}&\sum_{n=1}^{\infty}\left|\Phi_\eta^{n+1}(\xi,z)-\Phi_\eta^n(\xi,z)\right|\\ &\leq\tilde{C}_4\sum_{n=1}^{\infty}n\gamma_4^n\leq\frac{\tilde{C}_4}{(1-\gamma_4)^2}<\infty.\end{aligned}\tag{107}$$

Now, combining (95), (103), (105), we get

$$\begin{aligned}\left|\Phi_\eta^n(\xi,z')-\Phi_\eta^n(\xi,z'')\right|\leq&\sum_{k=1}^{n}\left|A_\eta^{k,n}(\xi,z')\right|+\sum_{k=1}^{n}\left|A_\eta^{k,n}(\xi,z'')\right|\\ &+\left|B_\eta^n(\xi,z')\right|+\left|B_\eta^n(\xi,z'')\right|\\ \leq&2\tilde{C}_2n\gamma_4^n+2\tilde{C}_3\gamma_4^n.\end{aligned}$$

Then, (102) implies

$$\begin{aligned}\left|\Phi_\eta^n(\xi',z')-\Phi_\eta^n(\xi'',z'')\right|\leq&\left|\Phi_\eta^n(\xi',z')-\Phi_\eta^n(\xi'',z')\right|\\ &+\left|\Phi_\eta^n(\xi'',z')-\Phi_\eta^n(\xi'',z'')\right|\\ \leq&\tilde{C}_2(2n+1)\gamma_4^n+2\tilde{C}_3\gamma_4^n\leq\tilde{C}_4n\gamma_4^n.\end{aligned}\tag{108}$$

Let $C_7 = \tilde{C}_4/(1-\gamma_4)^2$. Moreover, let

$$\phi_\eta(\xi,z)=\Phi_\eta^0(\xi,z)+\sum_{n=0}^{\infty}\left(\Phi_\eta^{n+1}(\xi,z)-\Phi_\eta^n(\xi,z)\right).$$

Then, due to (107), $\phi_\eta(\xi,z)$ is well-defined. Now, (106) implies

$$\begin{aligned}\left|\Phi_\eta^n(\xi,z)-\phi_\eta(\xi,z)\right|\leq&\sum_{k=n}^{\infty}\left|\Phi_\eta^{k+1}(\xi,z)-\Phi_\eta^k(\xi,z)\right|\\ \leq&\tilde{C}_4\sum_{k=n}^{\infty}k\gamma_4^k\leq C_7n\gamma_4^n.\end{aligned}\tag{109}$$

Consequently, (108) yields

$$\begin{aligned}|\phi_\eta(\xi',z')-\phi_\eta(\xi'',z'')|\leq&\left|\Phi_\eta^n(\xi',z')-\Phi_\eta^n(\xi'',z'')\right|\\ &+\left|\Phi_\eta^n(\xi',z')-\phi_\eta(\xi',z')\right|\\ &+\left|\Phi_\eta^n(\xi'',z'')-\phi_\eta(\xi'',z'')\right|\\ \leq&3C_7n\gamma_4^n\end{aligned}$$

Therefore, $\phi_\eta(\xi',z')=\phi_\eta(\xi'',z'')$ for any $\xi',\xi''\in V_{\delta_9}(\mathcal{P}(\mathcal{X}))$, $z',z''\in\mathcal{Z}$. Hence, there exists a function ϕ_η which maps $\eta\in\mathbb{C}^d$ to $\mathbb{C}$ and satisfies $\phi_\eta=\phi_\eta(\xi,z)$ for all

$\eta \in V_{\delta_9}(\Theta)$, $\xi \in V_{\delta_9}(\mathcal{P}(\mathcal{X}))$, $z \in \mathcal{Z}$. Then, using (109), we conclude that (97) holds for $\eta \in V_{\delta_9}(\Theta)$, $\xi \in V_{\delta_9}(\mathcal{P}(\mathcal{X}))$, $z \in \mathcal{Z}$. ☐

Lemma 7.2. *(i) Let Assumptions 2.1 – 2.3 and 2.5 hold. Then, integral*

$$\int \cdots \int \Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right) S(z_n, dz_{n+1}) \cdots S(z, dz_1) \tag{110}$$

is analytic in η for all $\eta \in V_{\delta_5}(\Theta)$, $\lambda \in \mathcal{P}(\mathcal{X})$, $z \in \mathcal{Z}$, $n \geq 1$ (δ_5 is specified in Lemmas 6.5, 6.6).

(ii) Let Assumptions 2.1 – 2.4 hold. Then, integral

$$\int \cdots \int \Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right) S_\eta(z_n, dz_{n+1}) \cdots S_\eta(z, dz_1) \tag{111}$$

is analytic in η for all $\eta \in V_{\delta_5}(\Theta)$, $\lambda \in \mathcal{P}(\mathcal{X})$, $z \in \mathcal{Z}$, $n \geq 1$.

Proof. Throughout the proof, the following notations is used. $\tilde{\phi}(z)$ is the function defined by $\tilde{\phi}(z) = \phi(y)$ for $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z = (y, x)$. η is any element of $V_{\delta_5}(\Theta)$, while λ is any element in $\mathcal{P}(\mathcal{X})$. $\{x_n\}_{n\geq 0}$, $\{y_n\}_{n\geq 0}$ are any sequences in $\mathcal{X}$, $\mathcal{Y}$ (respectively), while $\{z_n\}_{n\geq 0}$ is the sequence defined by $z_n = (y_n, x_n)$ for $n \geq 0$ (notice that $\{z_n\}_{n\geq 0}$ is any sequence in $\mathcal{Z}$). $n \geq 1$ is any integer.

Using Lemmas 6.1, 6.6, we conclude

$$\begin{aligned}
\Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right) =& \Phi_{\eta, y_{n+1}} \left(F_{\eta, \boldsymbol{y}}^{0:n}(\lambda)\right) \\
=& \log \left\langle R_{\eta, \boldsymbol{y}}^{n:n+1} \left(F_{\eta, \boldsymbol{y}}^{0:n}(\lambda)\right)\right\rangle \\
=& \log \left(\frac{\left\langle R_{\eta, \boldsymbol{y}}^{0:n+1}(\lambda)\right\rangle}{\left\langle R_{\eta, \boldsymbol{y}}^{0:n}(\lambda)\right\rangle} \right),
\end{aligned}$$

where $\boldsymbol{y} = \{y'_k\}_{k\geq 1}$ is any sequence in $\mathcal{Y}$ satisfying $y'_k = y_k$ for $1 \leq k \leq n+1$. Combining this with Lemmas 5.3, 6.3, 6.6, we deduce that $\Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right)$, $u_\eta^n(z_{0:n})$ are analytic in η for each $\eta \in V_{\delta_5}(\Theta)$. On the other side, due to Lemmas 5.3, 6.4, 6.6, we have

$$\left|\Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right)\right| \leq C_4 \tilde{\psi}(z_{n+1}), \tag{112}$$

$$\left|u_\eta^n(z_{0:n})\right| \leq K_n \prod_{k=1}^n \tilde{\phi}(z_k) \tag{113}$$

($\tilde{\psi}(z)$ is specified in (11)).

Owing to Assumption 2.5, we have

$$\begin{aligned}
&\int \cdots \int \tilde{\psi}(z_{n+1}) S(z_n, dz_{n+1}) \cdots S(z_0, dz_1) \\
=& \int \int \cdots \int (1 + \psi(y_{n+1})) Q(x_{n+1}, dy_{n+1}) \\
&\cdot P(x_n, dx_{n+1}) \cdots P(x_0, dx_1) \\
\leq& K + 1 < \infty
\end{aligned}$$

(notice that $z_k = (y_k, x_k)$). Consequently, Lemma A1.1 (see Appendix 1) and (112) imply that integral (110) is analytic in η for each $\eta \in V_{\delta_5}(\Theta)$.

Relying on (15), it is easy to show

$$\begin{aligned}
&\int \cdots \int \Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right) S_\eta(z_n, dz_{n+1}) \cdots S_\eta(z_0, dz_1) \\
=& \int \cdots \int \Phi_\eta \left(F_\eta^n(\lambda, z_{1:n}), z_{n+1}\right) u_\eta^{n+1}(z_{0:n+1}) \\
&\cdot (\nu \times \mu)(dz_{n+1}) \cdots (\nu \times \mu)(dz_1)
\end{aligned}$$

On the other side, due to Assumptions 2.2, 2.4, we have

$$\begin{aligned}
&\int \cdots \int \tilde{\psi}(z_{n+1}) \left(\prod_{k=1}^{n+1} \tilde{\phi}(z_k) \right) \\
&\cdot (\nu \times \mu)(dz_{n+1}) \cdots (\nu \times \mu)(dz_1) \\
=& \|\mu\|^{n+1} \left(\int (1 + \psi(y_{n+1})) \phi(y_{n+1}) \nu(dy_{n+1}) \right) \\
&\cdot \left(\prod_{k=1}^n \int \phi(y_k) \nu(dy_k) \right) \\
<& \infty
\end{aligned}$$

Consequently, Lemma A1.1 (see Appendix 1) and (112), (113) imply that integral (111) is analytic in η for $\eta \in V_{\delta_5}(\Theta)$. ☐

Proof of Theorem 2.1. Let $T_\eta(z, dz')$ be the kernel defined by $T_\eta(z, B) = S(z, B)$ for $\eta \in \mathbb{C}^d$, $z \in \mathcal{Z}$, $B \in \mathcal{B}(\mathcal{Z})$ ($S(z, dz')$ is specified in (13)). Moreover, let $T_\eta^n(z, dz')$, $\Phi_\eta^n(\lambda, z)$ have the same meaning as in (93). Then, owing to Lemma 7.2, $\Phi_\eta^n(\lambda, z)$ is analytic in η for each $\eta \in V_{\delta_5}(\Theta)$, $\lambda \in \mathcal{P}(\mathcal{X})$, $z \in \mathcal{Z}$, $n \geq 1$. On the other side, due to Lemma 5.1, kernel $T_\eta(z, dz')$ (defined here) satisfies Assumptions 7.1, 7.2. Combining this with Lemma 7.1, we deduce that there exist a function ϕ_η mapping $\eta \in \mathbb{C}^d$ to $\mathbb{C}$ and real numbers $\delta_9 \in (0, \delta_5]$, $\gamma_4 \in (0, 1)$, $C_7 \in [1, \infty)$ such that (97) holds for $\eta \in V_{\delta_9}(\Theta)$, $\lambda \in \mathcal{P}(\mathcal{X})$, $z \in \mathcal{Z}$, $n \geq 1$. Since the limit of uniformly convergent analytic functions is also analytic (see e.g., [28, Theorem 2.4.1]), ϕ_η is analytic in η for each $\eta \in V_{\delta_9}(\Theta)$.

In what follows in the proof, θ, λ, z are any elements of Θ, $\mathcal{P}(\mathcal{X})$, $\mathcal{Z}$ (respectively), while $n \geq 1$ is any integer. It is straightforward to verify

$$\Phi_\theta^n(\lambda, z) = E\left(\Phi_\theta\left(F_\theta^n(\lambda, Z_{1:n}), Z_{n+1}\right) \middle| Z_0 = z \right),$$

where $Z_n = (Y_n, X_n)$. Therefore, (94) yields

$$E\left(\log q_\theta^n(Y_{1:n}|\lambda)\right) = \sum_{k=0}^{n-1} E\left(\Phi_\theta^k(\lambda, Z_0)\right).$$

Then, Lemma 7.1 implies

$$\begin{aligned}
\left| E\left(\frac{1}{n} \log q_\theta(Y_{1:n}|\lambda)\right) - \phi_\theta \right| \leq& \frac{1}{n} \sum_{k=0}^{n-1} E\left|\Phi_\theta^k(\lambda, Z_0) - \phi_\theta\right| \\
\leq& \frac{C_7}{n} \sum_{k=0}^{n-1} \gamma_4^k \leq \frac{C_7}{n(1 - \gamma_4)}.
\end{aligned}$$

Consequently, there exists a function $l : \Theta \to \mathbb{R}$ with the properties specified in the statement of the theorem. ☐

Proof of Theorem 2.2. Let $T_\eta(z, dz')$ be the kernel defined by $T_\eta(z, B) = S_\eta(z, B)$ for $\eta \in \mathbb{C}^d$, $z \in \mathcal{Z}$, $B \in \mathcal{B}(\mathcal{Z})$ ($S_\eta(z, dz')$ is specified in (14)). Moreover, let $T_\eta^n(z, dz')$,

$\Phi_\eta^n(\lambda,z)$ have the same meaning as in (93). Then, due to Lemma 7.2, $\Phi_\eta^n(\lambda,z)$ is analytic in η for each $\eta\in V_{\delta_5}(\Theta)$, $\lambda\in\mathcal{P}(\mathcal{X})$, $z\in\mathcal{Z}$, $n\geq 1$. On the other side, Lemma 5.4 implies that Assumptions 7.1, 7.2 hold for kernel $T_\eta(z,dz')$ (defined here). Combining this with Lemma 7.1, we conclude that there exist a function ϕ_η mapping $\eta\in\mathbb{C}^d$ to $\mathbb{C}$ and real numbers $\delta_9\in(0,\delta_5]$, $\gamma_4\in(0,1)$, $C_7\in[1,\infty)$ such that (97) holds for $\eta\in V_{\delta_9}(\Theta)$, $\lambda\in\mathcal{P}(\mathcal{X})$, $z\in\mathcal{Z}$, $n\geq 1$. As the limit of uniformly convergent analytic functions is also analytic (see e.g., [28, Theorem 2.4.1]), ϕ_η is analytic in η for each $\eta\in V_{\delta_9}(\Theta)$.

In the rest of the proof, θ, λ, z are any elements of Θ, $\mathcal{P}(\mathcal{X})$, $\mathcal{Z}$ (respectively), while $n\geq 1$ is any integer. It is easy to show

$$\Phi_\theta^n(\lambda,z)=E\left(\Phi_\theta\left(F_\theta^n\left(\lambda,Z_{1:n}^{\theta,\lambda}\right),Z_{n+1}^{\theta,\lambda}\right)\middle|\,Z_0^{\theta,\lambda}=z\right),$$

where $Z_n^{\theta,\lambda}=\left(Y_n^{\theta,\lambda},X_n^{\theta,\lambda}\right)$. Then, (94) yields

$$E\left(\log q_\theta^n\left(Y_{1:n}^{\theta,\lambda}|\lambda\right)\right)=\sum_{k=0}^{n-1}E\left(\Phi_\theta^k\left(\lambda,Z_0^{\theta,\lambda}\right)\right).$$

Therefore, Lemma 7.1 implies

$$\begin{aligned}\left|E\left(\frac{1}{n}\log q_\theta\left(Y_{1:n}^{\theta,\lambda}|\lambda\right)\right)-\phi_\theta\right|\leq&\frac{1}{n}\sum_{k=0}^{n-1}E\left|\Phi_\theta^k\left(\lambda,Z_0^{\theta,\lambda}\right)-\phi_\theta\right|\\ \leq&\frac{C_7}{n}\sum_{k=0}^{n-1}\gamma_4^k\leq\frac{C_7}{n(1-\gamma_4)}.\end{aligned}$$

Consequently, there exists a function $h:\Theta\to\mathbb{R}$ with the properties specified in the statement of the theorem. □

VIII. Proof of Corollaries 3.1 – 4.2

Proof of Corollaries 3.1 and 3.2. Let $\tilde{\Theta}$ be any non-empty bounded open set satisfying $\text{cl}\tilde{\Theta}\subset\Theta$. As $\text{cl}\tilde{\Theta}$, $\mathcal{X}$ are compact sets, Assumption 3.2 and Lemma A1.2 (see Appendix 1) imply that there exist functions $\{\hat{a}_\eta^i(x)\}_{1\leq i\leq N_x}$, $\{\hat{b}_\eta^j(x)\}_{1\leq j\leq N_y}$ with the following properties:

(i) $\{\hat{a}_\eta^i(x)\}_{1\leq i\leq N_x}$, $\{\hat{b}_\eta^j(x)\}_{1\leq j\leq N_y}$ map $\eta\in\mathbb{C}^d$, $x\in\mathbb{C}^{d_x}$ to $\mathbb{C}$.

(ii) $\hat{a}_\theta^i(x)=a_\theta^i(x)$, $\hat{b}_\theta^j(x)=b_\theta^j(x)$ for $\theta\in\tilde{\Theta}$, $x\in\mathcal{X}$, $1\leq i\leq N_x$, $1\leq j\leq N_y$.

(iii) There exists a real number $\alpha_1\in(0,1)$ such that $\hat{a}_\eta^i(x)$, $\hat{b}_\eta^j(x)$ are analytic in (η,x) for $\eta\in V_{\alpha_1}(\tilde{\Theta})$, $x\in V_{\alpha_1}(\mathcal{X})$, $1\leq i\leq N_x$, $1\leq j\leq N_y$.

Owing to Assumption 3.2, $\{\hat{a}_\theta^i(x)\}_{1\leq i\leq N_x}$, $\{\hat{b}_\theta^j(x)\}_{1\leq j\leq N_y}$ are positive and uniformly bounded away from zero for $\theta\in\text{cl}\tilde{\Theta}$, $x\in\mathcal{X}$. Then, due to (iii), there exist real numbers $\alpha\in(0,\alpha_1)$, $\beta\in(0,1)$ such that

$$\text{Re}\left\{\hat{a}_\eta^i(x)\right\}\geq\beta,\quad|\hat{a}_\eta^i(x)|\leq\frac{1}{\beta},\tag{114}$$

$$\text{Re}\left\{\hat{b}_\eta^j(x)\right\}\geq\beta,\quad|\hat{b}_\eta^j(x)|\leq\frac{1}{\beta}\tag{115}$$

for $\eta\in V_\alpha(\tilde{\Theta})$, $x\in V_\alpha(\mathcal{X})$, $1\leq i\leq N_x$, $1\leq j\leq N_y$.

Let $\hat{p}_\eta(x'|x)$, $\hat{q}_\eta(y|x)$ be the functions defined by

$$\hat{p}_\eta(x'|x)=\sum_{i=1}^{N_x}\hat{a}_\eta^i(x)v_i(x'),\qquad\hat{q}_\eta(y|x)=\sum_{j=1}^{N_y}\hat{b}_\eta^j(x)w_j(y)$$

for $\eta\in\mathbb{C}^d$, $x,x'\in\mathcal{X}$, $y\in\mathcal{Y}$, while

$$\begin{aligned}r_\theta(y,x'|x)&=q_\theta(y|x')p_\theta(x'|x),\\ \hat{r}_\eta(y,x'|x)&=\hat{q}_\eta(y|x')\hat{p}_\eta(x'|x)\end{aligned}$$

for the same η,x,x',y and $\theta\in\Theta$. Then, owing to (ii), (iii), $\hat{r}_\eta(y,x'|x)$ is analytic in η for each $\eta\in V_\alpha(\tilde{\Theta})$, $x,x'\in\mathcal{X}$, $y\in\mathcal{Y}$. Due to the same reasons, $\hat{r}_\theta(y,x'|x)=r_\theta(y,x'|x)$ for the same x,x',y and $\theta\in\tilde{\Theta}$. On the other side, Assumption 3.3 and (114) imply

$$|\hat{p}_\eta(x'|x)|\geq\sum_{i=1}^{N_x}\text{Re}\left\{\hat{a}_\eta^i(x)\right\}v_i(x')\geq\beta\varepsilon N_x,\tag{116}$$

$$|\hat{p}_\eta(x'|x)|\leq\sum_{i=1}^{N_x}|\hat{a}_\eta^i(x)|v_i(x')\leq\frac{N_x}{\beta\varepsilon}\tag{117}$$

for $\eta\in V_\alpha(\tilde{\Theta})$, $x,x'\in\mathcal{X}$. Similarly, (115) yields

$$|\hat{q}_\eta(y|x)|\geq\sum_{j=1}^{N_y}\text{Re}\left\{\hat{b}_\eta^j(x)\right\}w_j(y)\geq\beta\sum_{j=1}^{N_y}w_j(y),\tag{118}$$

$$|\hat{q}_\eta(y|x)|\leq\sum_{j=1}^{N_y}|\hat{b}_\eta^j(x)|w_j(y)\leq\frac{1}{\beta}\sum_{j=1}^{N_y}w_j(y)\tag{119}$$

for the same η,x and $y\in\mathcal{Y}$.

Let $\tilde{C}_1=\beta^{-2}\varepsilon^{-1}N_x$, $\tilde{C}_2=\tilde{C}_1N_x$, $\gamma=\beta^4\varepsilon^2$. Moreover, let $\phi(y)$, $\psi(y)$ be the functions defined by

$$\phi(y)=\tilde{C}_1\sum_{j=1}^{N_y}w_j(y),\quad\psi(y)=\tilde{C}_2\left(1+\sum_{j=1}^{N_y}|\log w_j(y)|\right)$$

for $y\in\mathcal{Y}$. Then, combining (116) – (119), we get

$$\gamma\phi(y)\leq|\hat{r}_\eta(y,x'|x)|\leq\frac{\phi(y)}{\gamma}\tag{120}$$

for $\eta\in V_\alpha(\tilde{\Theta})$, $x,x'\in\mathcal{X}$, $y\in\mathcal{Y}$. We also get

$$\begin{aligned}\log\phi(y)\leq&\log(\tilde{C}_1N_x)+\max_{1\leq j\leq N_y}\log w_j(y)\\ \leq&\tilde{C}_1N_x\left(1+\sum_{j=1}^{N_y}|\log w_j(y)|\right),\\ \log\phi(y)\geq&\log(\tilde{C}_1N_x)+\min_{1\leq j\leq N_y}\log w_j(y)\\ \geq&-\tilde{C}_1N_x\left(1+\sum_{j=1}^{N_y}|\log w_j(y)|\right).\end{aligned}$$

Therefore, we have

$$|\log\phi(y)|\leq\psi(y).\tag{121}$$

Since $\int\phi(y)\nu(dy)=\tilde{C}_1N_y<\infty$, (120), (121) imply that Assumptions 2.1 – 2.3 follow from Assumptions 4.1 – 4.3 when Θ is restricted to $\tilde{\Theta}$ (i.e., when Θ is replaced with $\tilde{\Theta}$).

Owing to Assumption 3.5, we have

$$\begin{aligned}\int\psi(y)Q(x,dy)=&\tilde{C}_2+\sum_{j=1}^{N_y}|\log w_j(y)|Q(x,dy)\\ \leq&\tilde{C}_2+KN_y<\infty.\end{aligned}$$

Hence, Assumption 2.5 results from Assumption 3.5. Moreover, due to Assumption 3.4, we have

$$\int \psi(y)\phi(y)\nu(dy) = \tilde{C}_1\tilde{C}_2 \sum_{j,k=1}^{N_y} \int |\log w_j(y)| w_k(y)\nu(dy) + \tilde{C}_1\tilde{C}_2 N_y < \infty.$$

Thus, Assumption 2.4 results from Assumption 3.4.

Using Theorems 2.1, 2.2, we conclude that there exist functions $\tilde{l}, \tilde{h} : \tilde{\Theta} \to \mathbb{R}$ such that $\tilde{l}(\theta)$, $\tilde{h}(\theta)$ are real-analytic in θ and satisfy $\lim_{n\to\infty} l_n(\theta,\lambda) = \tilde{l}(\theta)$, $\lim_{n\to\infty} h_n(\theta,\lambda) = \tilde{h}(\theta)$ for each $\theta \in \tilde{\Theta}$, $\lambda \in \mathcal{P}(\mathcal{X})$ ($l_n(\theta,\lambda)$, $h_n(\theta,\lambda)$ have the same meaning as in (1)). Consequently, Corollaries 4.1, 4.2 hold (notice that Θ can be represented as the union $\Theta = \bigcup_{n=1}^{\infty} \tilde{\Theta}_n$, where $\{\tilde{\Theta}_n\}_{n\geq 1}$ are non-empty open balls satisfying $\text{cl}\tilde{\Theta}_n \subset \Theta$ for $n \geq 1$). $\square$

Proof of Corollaries 4.1 and 4.2. Let $\tilde{\Theta}$ be a non-empty bounded open set satisfying $\text{cl}\tilde{\Theta} \subset \Theta$. As $\text{cl}\tilde{\Theta}$, $\mathcal{X}$, $\mathcal{Y}$ are compact sets, Assumption 4.3 implies that there exists a real number $r \in [1,\infty)$ such that

$$\left\| B_\theta^{-1}(x)\left(x' - A_\theta(x)\right)\right\| \leq r, \quad \left\| D_\theta^{-1}(x)\left(y - C_\theta(x)\right)\right\| \leq r \tag{122}$$

for $\theta \in \text{cl}\tilde{\Theta}$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ (notice that $A_\theta(x)$, $B_\theta^{-1}(x)$, $C_\theta(x)$, $D_\theta^{-1}(x)$ are continuous in (θ, x)).

Let $\tilde{\mathcal{X}} = \{x \in \mathbb{R}^{d_x} : \|x\| \leq r\}$, $\tilde{\mathcal{Y}} = \{y \in \mathbb{R}^{d_y} : \|y\| \leq r\}$. Since $\text{cl}\tilde{\Theta}$, $\mathcal{X}$, $\tilde{\mathcal{X}}$, $\tilde{\mathcal{Y}}$ are compact sets, Assumptions 4.2, 4.3 and Lemma A1.2 (see Appendix 1) imply that there exist functions $\hat{A}_\eta(x)$, $\hat{B}_\eta(x)$, $\hat{C}_\eta(x)$, $\hat{D}_\eta(x)$ and $\hat{v}(x)$, $\hat{w}(y)$ with the following properties:

(i) $\hat{A}_\eta(x)$, $\hat{B}_\eta(x)$, $\hat{C}_\eta(x)$, $\hat{D}_\eta(x)$ map $\eta \in \mathbb{C}^d$, $x \in \mathbb{C}^{d_x}$ to $\mathbb{C}^{d_x}$, $\mathbb{C}^{d_x\times d_x}$, $\mathbb{C}^{d_y}$, $\mathbb{C}^{d_y\times d_y}$ (respectively), while $\hat{v}(x)$, $\hat{w}(y)$ map $x \in \mathbb{C}^{d_x}$, $y \in \mathbb{C}^{d_y}$ to $\mathbb{C}$.

(ii) $\hat{A}_\theta(x) = A_\theta(x)$, $\hat{B}_\theta(x) = B_\theta(x)$, $\hat{C}_\theta(x) = C_\theta(x)$, $\hat{D}_\theta(x) = D_\theta(x)$ for $\theta \in \tilde{\Theta}$, $x \in \mathcal{X}$, and $\hat{v}(x) = v(x)$, $\hat{w}(y) = w(y)$ for $x \in \tilde{\mathcal{X}}$, $y \in \tilde{\mathcal{Y}}$.

(iii) There exists a real number $\alpha_1 \in (0,1)$ such that $\hat{A}_\eta(x)$, $\hat{B}_\eta(x)$, $\hat{C}_\eta(x)$, $\hat{D}_\eta(x)$ are analytic in (η, x) for $\eta \in V_{\alpha_1}(\tilde{\Theta})$, $x \in V_{\alpha_1}(\mathcal{X})$.

(iv) There exists a real number $\alpha_2 \in (0,1)$ such that $\hat{v}(x)$, $\hat{w}(y)$ are analytic in x, y (respectively) for $x \in V_{\alpha_2}(\tilde{\mathcal{X}})$, $y \in V_{\alpha_2}(\tilde{\mathcal{Y}})$.

Owing to Assumption 4.3 and (iii), there exists a real number $\alpha_3 \in (0,\alpha_1)$ such that $\det\hat{B}_\eta(x) \neq 0$, $\det\hat{D}_\eta(x) \neq 0$ for $\eta \in V_{\alpha_3}(\tilde{\Theta})$, $x \in V_{\alpha_3}(\mathcal{X})$ (notice that $|\det\hat{B}_\theta(x)|$, $|\det\hat{D}_\theta(x)|$ are uniformly bounded away from zero for $\theta \in \text{cl}\tilde{\Theta}$, $x \in \mathcal{X}$). Therefore,

$$\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right), \qquad \hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)$$

are well-defined and analytic in (η, x, x', y) for $\eta \in V_{\alpha_3}(\tilde{\Theta})$, $x \in V_{\alpha_3}(\mathcal{X})$, $x' \in \mathbb{C}^{d_x}$, $y \in \mathbb{C}^{d_y}$. Consequently, there exists a real number $\alpha_4 \in (0,\alpha_3)$ such that

$$\begin{aligned} &\left\| \text{Re}\left\{ \hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right\}\right\| < r + \frac{\alpha_2}{2}, \\ &\left\| \text{Im}\left\{ \hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right\}\right\| < \frac{\alpha_2}{2}, \\ &\left\| \text{Re}\left\{ \hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right\}\right\| < r + \frac{\alpha_2}{2}, \\ &\left\| \text{Im}\left\{ \hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right\}\right\| < \frac{\alpha_2}{2} \end{aligned}$$

for $\eta \in V_{\alpha_4}(\tilde{\Theta})$, $x, x' \in V_{\alpha_4}(\mathcal{X})$, $y \in V_{\alpha_4}(\mathcal{Y})$ (notice that $\text{cl}\tilde{\Theta}$, $\mathcal{X}$, $\mathcal{Y}$ are compact sets and use (122)). Hence, we have

$$\begin{aligned} &\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right) \in V_{\alpha_2}(\tilde{\mathcal{X}}), \\ &\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right) \in V_{\alpha_2}(\tilde{\mathcal{Y}}) \end{aligned}$$

for the same η, x, x', y. Then, (iv) implies that

$$\hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right), \quad \hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right) \tag{123}$$

are analytic in (η, x, x', y) for $\eta \in V_{\alpha_4}(\tilde{\Theta})$, $x, x' \in V_{\alpha_4}(\mathcal{X})$, $y \in V_{\alpha_4}(\mathcal{Y})$. Combining this with Assumption 4.2, we deduce that there exist real numbers $\alpha \in (0,\alpha_4)$, $\beta \in (0,1)$ such that

$$\text{Re}\left\{\hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right)\right\} \geq \beta, \tag{124}$$

$$\left|\hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right)\right| \leq \frac{1}{\beta}, \tag{125}$$

$$\text{Re}\left\{\hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right)\right\} \geq \beta, \tag{126}$$

$$\left|\hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right)\right| \leq \frac{1}{\beta} \tag{127}$$

for $\eta \in V_\alpha(\tilde{\Theta})$, $x, x' \in V_\alpha(\mathcal{X})$, $y \in V_\alpha(\mathcal{Y})$ (notice that functions (123) are positive and uniformly bounded away from zero for $\eta \in \text{cl}\tilde{\Theta}$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$).

Owing to (124), (125), we have

$$\begin{aligned} &\left|\int_{\mathcal{X}} \hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right) dx'\right| \\ &\geq \int_{\mathcal{X}} \text{Re}\left\{\hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right)\right\} dx' \geq \beta \text{m}(\mathcal{X}) > 0, \end{aligned} \tag{128}$$

$$\begin{aligned} &\left|\int_{\mathcal{X}} \hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right) dx'\right| \\ &\leq \int_{\mathcal{X}} \left|\hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right)\right| dx' \leq \frac{\text{m}(\mathcal{X})}{\beta} \end{aligned} \tag{129}$$

for $\eta \in V_\alpha(\tilde{\Theta})$, $x \in V_\alpha(\mathcal{X})$, where $\text{m}(\mathcal{X})$ is the Lebesgue measure of $\mathcal{X}$. Similarly, due to (126), (127), we have

$$\begin{aligned} &\left|\int_{\mathcal{Y}} \hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right) dy\right| \\ &\geq \int_{\mathcal{Y}} \text{Re}\left\{\hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right)\right\} dy \geq \beta \text{m}(\mathcal{Y}) > 0, \end{aligned} \tag{130}$$

$$\begin{aligned} &\left|\int_{\mathcal{Y}} \hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right) dy\right| \\ &\leq \int_{\mathcal{Y}} \left|\hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right)\right| dy \leq \frac{\text{m}(\mathcal{Y})}{\beta} \end{aligned} \tag{131}$$

for the same η, x, where $\mathrm{m}(\mathcal{Y})$ is the Lebesgue measure of $\mathcal{Y}$. On the other side, Lemma A1.1 (see Appendix 1) and (125), (127) imply that

$$\int_{\mathcal{X}} \hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right) dx', \tag{132}$$

$$\int_{\mathcal{Y}} \hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right) dy \tag{133}$$

are analytic in (η, x) for $\eta \in V_\alpha(\tilde{\Theta})$, $x \in V_\alpha(\mathcal{X})$.

In the rest of the proof, the following notations is used. $\hat{p}_\eta(x'|x)$, $\hat{q}_\eta(y|x)$ are the functions defined by $\hat{p}_{\eta'}(x'|x) = 0$, $\hat{q}_{\eta'}(y|x) = 0$ and

$$\hat{p}_\eta(x'|x) = \frac{\hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x' - \hat{A}_\eta(x)\right)\right)}{\int_{\mathcal{X}} \hat{v}\left(\hat{B}_\eta^{-1}(x)\left(x'' - \hat{A}_\eta(x)\right)\right) dx''},$$

$$\hat{q}_\eta(y|x) = \frac{\hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y - \hat{C}_\eta(x)\right)\right)}{\int_{\mathcal{Y}} \hat{w}\left(\hat{D}_\eta^{-1}(x)\left(y' - \hat{C}_\eta(x)\right)\right) dy'}$$

for $\eta \in V_\alpha(\tilde{\Theta})$, $\eta' \in \mathbb{C}^d \setminus V_\alpha(\tilde{\Theta})$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, while $r_\theta(y, x'|x)$, $\hat{r}_\eta(y, x'|x)$ are the functions defined by

$$r_\theta(y, x'|x) = q_\theta(y|x')p_\theta(x'|x),$$
$$\hat{r}_\eta(y, x'|x) = \hat{q}_\eta(y|x')\hat{p}_\eta(x'|x)$$

for the same x, x', y and $\theta \in \Theta$, $\eta \in \mathbb{C}^d$.

As functions (123) and integrals (132), (133) are analytic in (η, x, x', y) for $\eta \in V_\alpha(\tilde{\Theta})$, $x, x' \in V_\alpha(\mathcal{X})$, $y \in V_\alpha(\mathcal{Y})$, it follows from (128), (130) that $\hat{r}_\eta(y, x'|x)$ is well-defined and analytic in η for the same η, x, x', y. Similarly, (128) – (131) imply that there exists a real number $\gamma \in (0, 1)$ such that $\gamma \leq |\hat{r}_\eta(y, x'|x)| \leq 1/\gamma$ for $\eta \in V_\alpha(\tilde{\Theta})$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. On the other side, (ii) yields $\hat{r}_\theta(y, x'|x) = r_\theta(y, x'|x)$ for the same x, x', y and $\theta \in \tilde{\Theta}$. Consequently, Assumptions 2.1 – 2.5 follow from Assumptions 4.1 – 4.4 when Θ is restricted to $\tilde{\Theta}$ (i.e., when Θ is replaced with $\tilde{\Theta}$). Then, using Theorems 2.1, 2.2, we conclude that there exist function $\tilde{l}, \tilde{h} : \tilde{\Theta} \to \mathbb{R}$ such that $\tilde{l}(\theta)$, $\tilde{h}(\theta)$ are real-analytic in θ and satisfy $\lim_{n\to\infty} l_n(\theta, \lambda) = \tilde{l}(\theta)$, $\lim_{n\to\infty} h_n(\theta, \lambda) = \tilde{h}(\theta)$ for $\theta \in \tilde{\Theta}$, $\lambda \in \mathcal{P}(\mathcal{X})$ ($l_n(\theta, \lambda)$, $h_n(\theta, \lambda)$ have the same meaning as in (1)). Consequently, Corollaries 4.1, 4.2 hold (notice that Θ can be represented as the union $\Theta = \bigcup_{n=1}^\infty \tilde{\Theta}_n$, where $\{\tilde{\Theta}_n\}_{n\geq 1}$ are non-empty open balls satisfying $\mathrm{cl}\tilde{\Theta}_n \subset \Theta$ for $n \geq 1$). □

Appendix 1

This section contains some auxiliary results which are relevant for the proof of Lemmas 5.2, 5.3, 6.3, 7.2 and Corollaries 3.1 – 4.2. Here, we rely on the following notations. $d_w \geq 1$ and $d_z \geq 1$ are integers, while A is a bounded convex set in $\mathbb{C}^{d_w}$. $F(w, z)$ is a function mapping $w \in \mathbb{C}^{d_w}$, $z \in \mathbb{R}^{d_z}$ to $\mathbb{C}$, while $\lambda(dz)$ is a measure on $\mathbb{R}^{d_z}$. $f(w)$ is the function defined by

$$f(w) = \int F(w, z)\lambda(dz)$$

for $w \in \mathbb{C}^{d_w}$.

Lemma A1.1. *Assume the following:*

(i) There exists a real number $\delta \in (0, 1)$ such that $F(w, z)$ is analytic in w for each $w \in V_\delta(A)$, $z \in \mathbb{R}^{d_z}$.

(ii) There exists a function $\phi : \mathbb{R}^{d_z} \to [1, \infty)$ such that $|F(w, z)| \leq \phi(z)$ for all $w \in V_\delta(A)$, $z \in \mathbb{R}^{d_z}$.

Then, we have

$$|F(w', z) - F(w'', z)| \leq \frac{d_w \phi(z)\|w' - w''\|}{\delta}$$

for all $w', w'' \in V_\delta(A)$, $z \in \mathbb{R}^{d_z}$. Moreover, if $\int \phi(z)\lambda(dz) < \infty$, then $f(w)$ is well-defined and analytic for all $w \in V_\delta(A)$.

Proof. Owing to Cauchy's inequality (see e.g., [28, Proposition 2.1.3]) and (i), (ii), we have

$$\|\nabla_x F(w, z)\| \leq \frac{d_w \phi(z)}{\delta} \tag{134}$$

for $w \in V_\delta(A)$, $z \in \mathbb{R}^{d_z}$. Consequently, we get

$$\begin{aligned} &|F(w', z) - F(w'', z)| \\ &= \left|\int_0^1 \left(\nabla_w F(tw' + (1-t)w'', z)\right)^T (w' - w'')dt\right| \\ &\leq \int_0^1 \|\nabla_w F(tw' + (1-t)w'', z)\| \, \|w' - w''\|dt \\ &\leq \frac{d_w \phi(z)\|w' - w''\|}{\delta} \end{aligned}$$

for $w', w'' \in V_\delta(A)$, $z \in \mathbb{R}^{d_z}$ (notice that $tw' + (1-t)w'' \in V_\delta(A)$ for $t \in [0, 1]$ since $V_\delta(A)$ is convex). On the other side, if $\int \phi(z)\lambda(dz) < \infty$, then the dominated convergence theorem and (134) imply that $f(w)$ is well-defined and differentiable for $w \in V_\delta(A)$. Consequently, $f(w)$ is analytic for $w \in V_\delta(A)$. □

In the rest of this appendix, we use the following notations. B is a compact set in $\mathbb{R}^{d_w}$, while $g(w)$ is a function mapping $w \in \mathbb{R}^{d_w}$ to $\mathbb{R}$ (d_w is specified at the beginning in the appendix).

Lemma A1.2. *Assume that there exists an open set C in $\mathbb{R}^{d_w}$ such that $B \subset C$ and $g(w)$ is real-analytic on C. Then, there exists a function $\hat{g}(w)$ with the following properties:*

(i) $\hat{g}(w)$ maps $w \in \mathbb{C}^{d_w}$ to $\mathbb{C}$.

(ii) $\hat{g}(w) = g(w)$ for all $w \in B$.

(iii) There exists a real number $\delta \in (0, 1)$ such that $\hat{g}(w)$ is analytic on $V_\delta(B)$.

Proof. First, we assume that B is connected (latter, this assumption is dropped). As $g(w)$ is real-analytic on C, $g(w)$ has an analytic continuation in an open vicinity of any point in C. Hence, there exist functions $\hat{g}(w, v)$, $\delta(v)$ with the following properties:

(iv) $\hat{g}(w, v)$, $\delta(v)$ map $w \in \mathbb{C}^{d_w}$, $v \in C$ to $\mathbb{C}$, $(0, 1)$ (respectively).

(v) $\hat{g}(w, v) = g(w)$ for $w \in V_{\delta(v)}(v) \cap \mathbb{R}^{d_w}$, $v \in C$.

(vi) $\hat{g}(w, v)$ is analytic in w for $w \in V_{\delta(v)}(v)$, $v \in C$.

Since B is compact, there exist an integer $M \geq 1$ and points $\{v_i\}_{1\leq i\leq M}$ such that $v_i \in B$ for $1 \leq i \leq M$ and $B \subset \bigcup_{i=1}^M V_{\delta(v_i)}(v_i)$. Let $\hat{g}_i(w) = \hat{g}(w, v_i)$, $V_i = V_{\delta(v_i)}(v_i)$ for $w \in \mathbb{C}^{d_w}$, $1 \leq i \leq M$. As B is connected, for each $1 \leq i \leq M$, there exists $1 \leq j \leq M$, $j \neq i$ such that $V_i \cap V_j \cap \mathbb{R}^{d_w} \neq \emptyset$.

On the other side, if $V_i \cap V_j \cap \mathbb{R}^{d_w} \neq \emptyset$, then $V_i \cap V_j \cap \mathbb{R}^{d_w}$ is a non-empty open set and $\hat{g}_i(w) = \hat{g}_j(w) = g(w)$ for $w \in V_i \cap V_j \cap \mathbb{R}^{d_w}$. Then, by the uniqueness of analytic continuation (see e.g., [15, Corollary 1.2.6]), for each $1 \leq i \leq M$, there exist $1 \leq j \leq M$, $j \neq i$ and a function $\hat{g}_{ij}(w)$ with the following properties:

(vii) $\hat{g}_{ij}(w)$ maps $w \in \mathbb{C}^{d_w}$ to $\mathbb{C}$.

(viii) $\hat{g}_{ij}(w)$ is analytic on $V_i \cup V_j$.

(ix) $\hat{g}_{ij}(w) = \hat{g}_i(w)$ for $w \in V_i$ and $\hat{g}_{ij}(w) = \hat{g}_j(w)$ for $w \in V_j$.

Following these arguments, we conclude that there exists a function $\hat{g}(w)$ with the following properties:

(x) $\hat{g}(w)$ maps $w \in \mathbb{C}^{d_w}$ to $\mathbb{C}$.

(xi) $\hat{g}(w)$ is analytic on $\bigcup_{i=1}^{M} V_i$.

(xii) $\hat{g}(w) = \hat{g}_i(w)$ for $w \in V_i$, $1 \leq i \leq M$.

Now, we drop the assumption that B is connected (i.e., B is any compact set in $\mathbb{R}^{d_w}$). Since B is compact, there exist an integer $N \geq 1$ and open sets $\{W_i\}_{1 \leq i \leq N}$ in $\mathbb{R}^{d_w}$ such that $W_i \subseteq C$, $B \cap W_i \neq \emptyset$, $W_i \cap W_j = \emptyset$ for $1 \leq i, j \leq N$, $i \neq j$ and $B \subset \bigcup_{i=1}^{N} W_i$. Let $B_i = B \cap W_i$ for $1 \leq i \leq N$. Hence, $\{B_i\}_{1 \leq i \leq N}$ are connected components of B, and thus, $\{B_i\}_{1 \leq i \leq N}$ are compact and disjoint. Then, according to what has already been shown, there exist open sets $\{U_i\}_{1 \leq i \leq N}$ in $\mathbb{C}^{d_w}$ and functions $\{\hat{g}_i(w)\}_{1 \leq i \leq N}$ with the following properties:

(xiii) $B_i \subset U_i$, $U_i \cap U_j = \emptyset$ for $1 \leq i, j \leq N$, $i \neq j$.

(xiv) $\hat{g}_i(w)$ maps $w \in \mathbb{C}^{d_w}$ to $\mathbb{C}$ for $1 \leq i \leq N$.

(xv) $\hat{g}_i(w) = g(w)$ for $w \in B_i$, $1 \leq i \leq N$.

(xvi) $\hat{g}_i(w)$ is analytic on U_i for $1 \leq i \leq N$.

Let $\hat{g}(w)$ be the function defined by $\hat{g}(w) = \hat{g}_i(w)$ for $w \in U_i$, $1 \leq i \leq N$ and $\hat{g}(w) = 0$ for $w \notin \bigcup_{i=1}^{N} U_i$. Due to (xiii), $\hat{g}(w)$ is well-defined. As B is compact and $B \subset \bigcup_{i=1}^{N} U_i$ (owing to (xiii)), there exists a real number $\delta \in (0,1)$ such that $B \subset V_\delta(B) \subset \bigcup_{i=1}^{N} U_i$. Then, (xv), (xvi) imply that $\hat{g}(w)$ is analytic on $V_\delta(B)$ and satisfies $\hat{g}(w) = g(w)$ for $w \in B$. ☐

APPENDIX 2

In this section, we show how Theorem 2.2 can be applied to finite-state hidden Markov models. We also provide a link between Theorem 2.2 and the results of [12]. Here, we assume that $\mathcal{X}$ has a finite number of elements. We also assume $\mathcal{X} = \{1, \dots, N\}$ and $\mu(x) = 1$ for each $x \in \mathcal{X}$ (in this case, $p_\theta(x'|x)$ is the conditional probability of $X_{n+1}^{\theta,\lambda} = x'$ given $X_n^{\theta,\lambda} = x$). Further to this, we introduce the following assumptions.

Assumption A2.1. *$p_\theta(x'|x)$ and $q_\theta(y|x)$ are real-analytic in θ for each $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Moreover, $p_\theta(x'|x)$ and $q_\theta(y|x)$ have complex-valued continuations $\hat{p}_\eta(x'|x)$ and $\hat{q}_\eta(y|x)$ with the following properties:*

(i) $\hat{p}_\eta(x'|x)$ and $\hat{q}_\eta(y|x)$ map $\eta \in \mathbb{C}^d$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ to $\mathbb{C}$.

(ii) $\hat{p}_\theta(x'|x) = p_\theta(x'|x)$ and $\hat{q}_\theta(y|x) = q_\theta(y|x)$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

(iii) There exists a real number $\delta \in (0,1)$ such that $\hat{p}_\eta(x'|x)$ and $\hat{q}_\eta(y|x)$ are analytic in η for each $\eta \in V_\delta(\Theta)$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

(iv) There exists a real number $\varepsilon \in (0,1)$ such that $\varepsilon \leq |\hat{p}_\eta(x'|x)| \leq 1/\varepsilon$ for all $\eta \in V_\delta(\Theta)$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

Assumption A2.2. *There exists a real number $\alpha \in (0,1)$ and a vector $\hat{\theta} \in \Theta$ with the following properties:*

(i) $q_{\hat{\theta}}(y|x) \neq 0$, $q_\theta(y|x)/q_{\hat{\theta}}(y|x) \geq \alpha$ and $|\hat{q}_\eta(y|x)/q_{\hat{\theta}}(y|x)| \leq 1/\alpha$ for all $\theta \in \Theta$, $\eta \in V_\delta(\Theta)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

(ii) $\int |\log q_{\hat{\theta}}(y|x')| \, q_{\hat{\theta}}(y|x) \nu(dy) < \infty$ for all $x, x' \in \mathcal{X}$.

Assumption A2.3. *There exists a real number $\beta \in (0,1)$, a vector $\hat{x} \in \mathcal{X}$ and functions $\tilde{\phi}, \tilde{\psi} : \mathcal{Y} \to (0, \infty)$ with the following properties:*

(i) $\hat{q}_\eta(y|\hat{x}) \neq 0$, $|\hat{q}_\eta(y|x)/\hat{q}_\eta(y|\hat{x})| \leq 1/\beta$ for all $\eta \in V_\delta(\Theta)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

(ii) $|\hat{q}_\eta(y|\hat{x})| \leq \tilde{\phi}(y)$ and $|\log |\hat{q}_\eta(y|\hat{x})|| \leq \tilde{\psi}(y)$ for all $\eta \in V_\delta(\Theta)$, $y \in \mathcal{Y}$.

(iii) $\int \tilde{\phi}(y) \nu(dy) < \infty$ and $\int \tilde{\psi}(y) \tilde{\phi}(y) \nu(dy) < \infty$.

Assumptions A2.1 – A2.3 are a particular case of Assumptions 2.1 – 2.4 (see Corollary A2.1 and its proof). At the same time, Assumptions A2.1 – A2.3 include, as a special case, all conditions which the results of [12] are based on.[5] Further to this, Assumptions A2.1 – A2.3 considerably simplify the conditions adopted in [12].

Corollary A2.1. *Let Assumption A2.1 and one of Assumptions A2.2, A2.3 hold. Then, all conclusions of Theorem 2.2 are true.*

Proof. It is sufficient to show that Assumptions 2.1 – 2.4 follow from Assumption A2.1 and one of Assumptions A2.2, A2.3.

(i) In this part of the proof, we demonstrate that Assumption 2.1 holds under Assumption A2.1. Let $\lambda_\theta(dx|y)$ be the measure on $\mathcal{X}$ defined by

$$\lambda_\theta(B|y) = \sum_{x \in \mathcal{X}} q_\theta(y|x) I_B(x) \mu(x)$$

for $\theta \in \Theta$, $y \in \mathcal{Y}$, $B \subseteq \mathcal{X}$. Then, Assumption A2.1 implies

$$\begin{aligned} \sum_{x' \in B} r_\theta(y, x'|x) I_B(x') \mu(x') \geq& \varepsilon \sum_{x' \in B} q_\theta(y|x') I_B(x') \mu(x') \\ =& \lambda_\theta(B|y), \\ \sum_{x' \in B} r_\theta(y, x'|x) I_B(x') \mu(x') \leq& \frac{1}{\varepsilon} \sum_{x' \in B} q_\theta(y|x') I_B(x') \mu(x') \\ =& \frac{\lambda_\theta(B|y)}{\varepsilon} \end{aligned}$$

for the same θ, y, B and $x \in \mathcal{X}$. Hence, Assumption 2.1 holds.

(ii) In the next part of the proof, we show that Assumptions 2.2 – 2.4 follow from Assumptions A2.1, A2.2. Let $\tilde{C}_1 = \alpha^{-1} \varepsilon^{-1}$, $\gamma = \alpha^2 \varepsilon^2$, while $\hat{r}_\eta(y, x'|x)$ and $\varphi_\eta(y)$ are the functions defined by

$$\hat{r}_\eta(y, x'|x) = \hat{q}_\eta(y|x') \hat{p}_\eta(x'|x), \quad \varphi_\eta(y) = \tilde{C}_1 \sum_{x \in \mathcal{X}} q_{\hat{\theta}}(y|x)$$

[5] Notice that Assumptions A2.1 and A2.2 follow (respectively) from [12, Conditions (a), (c.i)] and [12, Conditions (b), (c.iii), Equation (11)]. Notice also that Assumption A2.3 results from one of [12, Conditions (c.iii), (d.i), Equation (7)] and [12, Conditions (c.ii), (c.iii), (d.i), Equation (8)].

for $\eta \in \mathbb{C}^d$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Then, due to Assumption A2.1, $\hat{r}_\eta(y, x'|x)$ and $\varphi_\eta(y)$ are analytic in η for each $\eta \in V_\delta(\Theta)$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ (notice that $\varphi_\eta(y)$ is constant in η). On the other side, Assumptions A2.1, A2.2 yield $\varphi_\eta(y) \neq 0$ and

$$|\hat{r}_\eta(y, x'|x)| \leq \frac{|\hat{q}_\eta(y|x')|}{\varepsilon} \leq \frac{q_{\hat{\theta}}(y|x')|}{\alpha\varepsilon} = |\varphi_\eta(y)| \tag{135}$$

for the same η, x, x', y. Assumptions A2.1, A2.2 also imply

$$\sum_{x' \in \mathcal{X}} r_\theta(y, x'|x)\mu(x') \geq \varepsilon \sum_{x' \in \mathcal{X}} q_\theta(y|x') \geq \alpha\varepsilon \sum_{x' \in \mathcal{X}} q_{\hat{\theta}}(y|x') = \gamma|\varphi_\theta(y)| \tag{136}$$

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

Let $\tilde{C}_2 = \tilde{C}_1 N$, while $\phi(y)$ and $\psi(y)$ are the functions defined by

$$\phi(y) = \tilde{C}_1 \sum_{x \in \mathcal{X}} q_{\hat{\theta}}(y|x),$$
$$\psi(y) = \tilde{C}_2 \left(1 + \sum_{x \in \mathcal{X}} |\log q_{\hat{\theta}}(y|x)|\right)$$

for $y \in \mathcal{Y}$. Then, due to Part (ii) of Assumption A2.2, we have $\int \phi(y)\nu(dy) = \tilde{C}_1 N < \infty$ and

$$\int \psi(y)\phi(y)\nu(dy) = \tilde{C}_1\tilde{C}_2 \sum_{x, x' \in \mathcal{X}} \int |\log q_{\hat{\theta}}(y|x')| q_{\hat{\theta}}(y|x)\nu(dy) + \tilde{C}_1\tilde{C}_2 N < \infty. \tag{137}$$

We also have

$$\log|\varphi_\eta(y)| \leq \log(\tilde{C}_1 N) + \max_{x \in \mathcal{X}} \log q_{\hat{\theta}}(y|x) \leq \tilde{C}_1 N \left(1 + \sum_{x \in \mathcal{X}} |\log q_{\hat{\theta}}(y|x)|\right),$$
$$\log|\varphi_\eta(y)| \geq \log(\tilde{C}_1 N) + \min_{x \in \mathcal{X}} \log q_{\hat{\theta}}(y|x) \geq -\tilde{C}_1 N \left(1 + \sum_{x \in \mathcal{X}} |\log q_{\hat{\theta}}(y|x)|\right)$$

for $\eta \in V_\delta(\Theta)$, $y \in \mathcal{Y}$. Consequently, we get

$$|\varphi_\eta(y)| \leq \phi(y), \qquad |\log|\varphi_\eta(y)|| \leq \psi(y) \tag{138}$$

for the same η, y. Then, using (135) – (138), we conclude that Assumptions 2.2 – 2.4 hold.

(iii) In this part of the proof, we show that Assumptions 2.2 – 2.4 follow from Assumptions A2.1, A2.3. Let $\tilde{C} = \beta^{-1}\varepsilon^{-1}$, $\gamma = \beta\varepsilon^2$, while $\hat{r}_\eta(y, x'|x)$ and $\varphi_\eta(y)$ are the functions defined by

$$\hat{r}_\eta(y, x'|x) = \hat{q}_\eta(y|x')\hat{p}_\eta(x'|x), \quad \varphi_\eta(y) = \tilde{C}\hat{q}_\eta(y|\hat{x})$$

for $\eta \in \mathbb{C}^d$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Then, due to Assumption A2.1, $\hat{r}_\eta(y, x'|x)$ and $\varphi_\eta(y)$ are analytic in η for each $\eta \in V_\delta(\Theta)$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. On the other side, Assumptions A2.1, A2.3 yield $\varphi_\eta(y) \neq 0$ and

$$|\hat{r}_\eta(y, x'|x)| \leq \frac{|\hat{q}_\eta(y|x')|}{\varepsilon} \leq \frac{|\hat{q}_\eta(y|\hat{x})|}{\beta\varepsilon} = |\varphi_\eta(y)| \tag{139}$$

for η, x, x', y. Assumptions A2.1, A2.3 also imply

$$\sum_{x' \in \mathcal{X}} r_\theta(y, x'|x)\mu(x') \geq q_\theta(y|\hat{x})p_\theta(\hat{x}|x) \geq \varepsilon q_\theta(y|\hat{x}) = \gamma|\varphi_\theta(y)| \tag{140}$$

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

Let $\phi(y)$ and $\psi(y)$ be the functions defined by

$$\phi(y) = \tilde{C}\tilde{\phi}(y), \quad \psi(y) = \tilde{C}\big(1 + \tilde{\psi}(y)\big)$$

for $y \in \mathcal{Y}$. Then, Assumption A2.3 yields

$$|\varphi_\eta(y)| = \tilde{C}|\hat{q}_\eta(y|\hat{x})| \leq \tilde{C}\tilde{\phi}(y) = \phi(y) \tag{141}$$

for $\eta \in V_\delta(\Theta)$, $y \in \mathcal{Y}$. Assumption A2.3 also implies

$$|\log|\varphi_\eta(y)|| \leq \log\tilde{C} + |\log|\hat{q}_\eta(y|\hat{x})|| \leq \tilde{C}(1 + \tilde{\psi}(y)) = \psi(y) \tag{142}$$

for the same η, y. Then, using Part (iii) of Assumption A2.3, and (139) – (142), we conclude that Assumptions 2.2 – 2.4 hold. □

In rest of the section, we explain how Theorem 2.2 can further be extended in the context of finite-state hidden Markov models. Here, we rely on the following notations. $\mathcal{P}^N$ is the set of N-dimensional probability vectors, while e is the N-dimensional vector whose all elements are one. For $\theta \in \Theta$, $y \in \mathcal{Y}$, $R_\theta(y)$ is the $N \times N$ matrix whose (x', x)-entry is $r_\theta(y, x'|x)$, where $r_\theta(y, x'|x)$ has the same meaning as in Section II. $G_\theta(\lambda, y)$ and $h_\theta(\lambda, y)$ are the functions defined by

$$G_\theta(\lambda, y) = \frac{R_\theta(y)\lambda}{e^T R_\theta(y)\lambda}, \qquad h_\theta(\lambda, y) = \log\left(e^T R_\theta(y)\lambda\right)$$

for $\theta \in \Theta$, $\lambda \in \mathcal{P}^N$, $y \in \mathcal{Y}$. Regarding functions $r_\theta(y, x'|x)$, $G_\theta(\lambda, y)$ and $h_\theta(\lambda, y)$, we assume the following.

Assumption A2.4. *There exist a real number $\varepsilon \in (0, 1)$ and a function $s_\theta(y, x)$ mapping $\theta \in \Theta$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ to $[0, \infty)$ such that*

$$\varepsilon s_\theta(y, x') \leq r_\theta(y, x'|x) \leq \frac{s_\theta(y, x')}{\varepsilon}$$

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

Assumption A2.5. *$G_\theta(\lambda, y)$ and $h_\theta(\lambda, y)$ are real-analytic in (θ, λ) for all $\theta \in \Theta$, $\lambda \in \mathcal{P}^N$, $y \in \mathcal{Y}$. Moreover, $G_\theta(\lambda, y)$ and $h_\theta(\lambda, y)$ have complex-valued continuations $\hat{G}_\eta(\xi, y)$ and $\hat{h}_\eta(\xi, y)$ with the following properties:*

(i) $\hat{G}_\eta(\xi, y)$ and $\hat{h}_\eta(\xi, y)$ map $\eta \in \mathbb{C}^d$, $\xi \in \mathbb{C}^N$, $y \in \mathcal{Y}$ to $\mathbb{C}^N$ and $\mathbb{C}$ (respectively).

(ii) $\hat{G}_\theta(\lambda, y) = G_\theta(\lambda, y)$ and $\hat{h}_\theta(\lambda, y) = h_\theta(\lambda, y)$ for all $\theta \in \Theta$, $\lambda \in \mathcal{P}^N$, $y \in \mathcal{Y}$.

(iii) There exists a real number $\delta \in (0, 1)$ such that $\hat{G}_\eta(\xi, y)$ and $\hat{h}_\eta(\xi, y)$ are analytic in (η, ξ) for each $\eta \in V_\delta(\Theta)$, $\xi \in V_\delta(\mathcal{P}^N)$, $y \in \mathcal{Y}$.

(iv) There exist a real number $K \in [1, \infty)$ and a function $\tilde{\psi} : \mathcal{Y} \to [1, \infty)$ such that $\int \exp\big(\tilde{\psi}(y)\big)\tilde{\psi}(y)\nu(dy) < \infty$ and

$$\|\hat{G}_\eta(\xi, y)\| \leq K, \qquad |\hat{h}_\eta(\xi, y)| \leq \tilde{\psi}(y)$$

for all $\eta \in V_\delta(\Theta)$, $\xi \in V_\delta(\mathcal{P}^N)$, $y \in \mathcal{Y}$.

Assumption A2.4 corresponds to the stability of the hidden Markov model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$ and its optimal filter, while Assumption A2.5 is related to the parameterization of the model $\{(X_n^{\theta,\lambda}, Y_n^{\theta,\lambda})\}_{n\geq 0}$. Assumptions A2.4 and A2.5 are the same as the (corresponding) assumptions adopted in [23]. Further to this, Assumptions A2.4 and A2.5 include, as a particular case, all conditions which the results of [12] are based on.

Theorem A2.1. *Let Assumptions A2.4 and A2.5 hold. Then, all conclusions of Theorem 2.2 are true.*

Proof. Let e_i be the i-th standard unit vector in $\mathbb{R}^N$, where $1 \leq i \leq N$. Moreover, let $\hat{r}_\eta(y, x'|x)$ be the function defined by

$$\hat{r}_\eta(y, x'|x) = e_{x'}^T \hat{G}_\eta(e_x, y) \exp\left(\hat{h}_\eta(e_x, y)\right)$$

for $\eta \in \mathbb{C}^d$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, while $\varphi_\eta(y)$, $\phi(y)$ and $\psi(y)$ are the functions defined by

$$\varphi_\eta(y) = \phi(y) = K \exp\left(\tilde{\psi}(y)\right), \qquad \psi(y) = 2K\tilde{\psi}(y)$$

for the same η, y. Then, it is straightforward to demonstrate that Assumptions 2.1, 2.2 and 2.4 hold.

Let $\hat{s}_\eta(x)$ be the function defined by

$$\hat{s}_\eta(x) = \sum_{x' \in \mathcal{X}} \int \hat{r}_\eta(y, x'|x)\nu(dy)$$

for $\eta \in \mathbb{C}^d$, $x \in \mathcal{X}$, while $\tilde{r}_\eta(y, x'|x)$ is the function be defined by

$$\tilde{r}_\eta(y, x'|x) = \begin{cases} \hat{r}_\eta(y, x'|x)/\hat{s}_\eta(x), & \text{if } \hat{s}_\eta(x) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

for the same η, x and $x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Moreover, let $T_\eta(z, B)$ be the kernel defined by

$$T_\eta(z, B) = \sum_{x' \in \mathcal{X}} \int I_B(y', x')\tilde{r}_\eta(y', x'|x)\nu(dy')$$

for $\eta \in \mathbb{C}^d$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, a Borel-set $B \subseteq \mathcal{Y} \times \mathcal{X}$ and $z = (y, x)$. Since Assumptions 2.1, 2.2 and 2.4 hold, Lemma 5.4 implies that Assumptions 7.1, 7.2 hold, too.

Let $\Phi_\eta(\xi, z)$ be the function defined by

$$\Phi_\eta(\xi, z) = \hat{h}_\eta(\xi, y)$$

for $\eta \in \mathbb{C}^d$, $\xi \in C^N$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z = (y, x)$, while $\{F_{\eta,\boldsymbol{y}}^{m:n}(\xi)\}_{n\geq m\geq 0}$ are the functions recursively defined by $F_{\eta,\boldsymbol{y}}^{m:m}(\xi) = \xi$ and

$$F_{\eta,\boldsymbol{y}}^{m:n+1}(\xi) = \hat{G}_\eta\left(F_{\eta,\boldsymbol{y}}^{m:n}(\xi), y_{n+1}\right)$$

for the same η, ξ and a sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in $\mathcal{Y}$. Then, due to Lemma A1.1, the conclusions of Lemma 6.4 hold. On the other side, owing to [23, Lemma 3], the conclusions of Lemma 6.6 also hold (provided that elements of $\mathbb{C}^N$ are interpreted as complex measures on $\mathcal{X}$).

Combining Assumptions 7.1, 7.2 and the conclusions of Lemmas 6.4, 6.6, we get the conclusions of Lemmas 7.1, 7.2 (notice that Assumptions 7.1, 7.2 and the conclusions of Lemmas 6.4, 6.6 are sufficient for the conclusions of Lemmas 7.1, 7.2 to hold) Then, as a direct consequence of the conclusions of Lemmas 7.1, 7.2, we get the conclusions of Theorem 2.2. □

REFERENCES

[1] L. E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains,* Annals of Mathematical Statistics, 37 (1966), pp. 15541563.

[2] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models,* Springer-Verlag, 2005.

[3] D. Crisan and B. Rozovskii, *The Oxford Handbook of Nonlinear Filtering,* Oxford University Press, 2011.

[4] 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.

[5] 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.

[6] R. Douc, E. Moulines, and D. S. Stoffer, *Nonlinear Time Series: Theory, Methods and Applications with R Examples,* CRC Press, 2014.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory,* 2nd Edition, Wiley, 2006.

[8] Y. Ephraim and N. Merhav, *Hidden Markov Models,* IEEE Transactions on Information Theory, 48 (2008), pp. 1518 – 1569.

[9] R. M. Gray, *Entropy and Information Theory,* 2nd Edition, Springer-Verlag, 2011.

[10] G. Han and B. Marcus, *Analyticity of entropy rate of hidden Markov chains,* IEEE Transactions on Information Theory, 52 (2006), pp. 5251 – 5266.

[11] G. Han and B. Marcus, *Derivatives of entropy rate in special families of hidden Markov chains,* IEEE Transactions on Information Theory, 53 (2007), pp. 2642 – 2652.

[12] G. Han and B. Marcus, *Analyticity of entropy rate of hidden Markov chains with continuous alphabet,* IEEE Transactions on Information Theory, 61 (2015), pp. 3013 – 3028.

[13] G. Han, *A randomized algorithm for the capacity of finite-state channels,* IEEE Transactions on Information Theory, 61 (2015), pp. 3651 – 3669.

[14] T. Holliday, A. Goldsmith, and P. Glynn, *Capacity of finite state channels based on Lyapunov exponents of random matrices,* IEEE Transaction on Information Theory, 52 (2006), pp. 3509 – 3532.

[15] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions,* Birikhäuser, 2002.

[16] 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.

[17] 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.

[18] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability,* 2nd Edition, Cambridge University Press, 2009.

[19] E. Ordentlich and T. Weissman, *On the optimality of symbol-by-symbol filtering and denoising,* IEEE Transactions on Information Theory, 52 (2006), pp. 19 – 40.

[20] Y. Peres, *Domains of analytic continuation for the top Lyapunov exponent,* Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 28 (1992), pp. 131 – 148.

[21] 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.

[22] A. Schönhuth, *On analytic properties of entropy rate,* IEEE Transactions on Information Theory 55 (2009), pp. 2119 – 2127.

[23] 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.

[24] 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.

[25] V. B. Tadić and A. Doucet, *Asymptotic bias of stochastic gradient search,* Annals of Applied Probability, 27 (2017), pp. 3255–3304.

[26] 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.

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

[28] J. L. Taylor, *Several Complex Variables with Connections to Algebraic Geometry and Lie Groups,* American Mathematical Society, 2002.