Analyticity, Convergence and Convergence Rate of Recursive Maximum Likelihood Estimation in Hidden Markov Models

Vladislav B. Tadić[1]

Abstract

This paper considers the asymptotic behavior of the recursive maximum likelihood estimation in hidden Markov models. The paper is focused on the analytic properties of the asymptotic log-likelihood and on the point-convergence and convergence rate of the recursive maximum likelihood estimator. Using the principle of analytical continuation, the analyticity of the asymptotic log-likelihood is shown for analytically parameterized hidden Markov models. Relying on this fact and some results from differential geometry (Lojasiewicz inequality), the almost sure point-convergence of the recursive maximum likelihood algorithm is demonstrated, and relatively tight bounds on the convergence rate are derived. As opposed to the existing result on the asymptotic behavior of maximum likelihood estimation in hidden Markov models, the results of this paper are obtained without assuming that the log-likelihood function has an isolated maximum at which the Hessian is strictly negative definite.

Index Terms

Hidden Markov models, maximum likelihood estimation, recursive identification, analyticity, Lojasiewicz inequality, point-convergence, convergence rate.

I. Introduction

Hidden Markov models are a broad class of stochastic processes capable of modeling complex correlated data and large-scale dynamical systems. These processes consist of two components: states and observations. The states are unobservable and form a Markov chain. The observations are independent conditionally on the states and provide only available information about the state dynamics. Hidden Markov models have been formulated in the seminal paper [1], and over last few decades, they have found a wide range of applications in diverse areas such as acoustics and signal processing, image analysis and computer vision, automatic control and robotics, economics and finance, computation biology and bioinformatics. Due to their practical relevance, these models have extensively been studied in a large number of papers and books (see e.g., [8], [12] and references cited therein).

[1] Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom. (v.b.tadic@bristol.ac.uk).

Besides the estimation of states given available observations (also known as filtering), the identification of model parameters are probably the most important problem associated with hidden Markov models. This problem can be described as the estimation (or approximation) of the state transition probabilities and the observations conditional distributions given available observations. The identification of hidden Markov models have been considered in numerous papers and several methods and algorithms have been developed (see [8, Part II], [12] and references cited therein). Among them, the methods based on the maximum likelihood principle are probably one of the most important and popular. Their various asymptotic properties (asymptotic consistency, asymptotic normality, convergence rate) have been analyzed in a number of papers (see [1], [5], [6], [10], [11], [19] – [22], [24], [28], [33], [34]; see also [8, Chapter 12], [12] and references cited therein). Although the existing results provide an excellent insight into the asymptotic behavior of maximum likelihood estimators for hidden Markov models, they all crucially rely on the assumption that the log-likelihood function has a strong maximum, i.e., an isolated maximum at which the Hessian is strictly negative definite. As the log-likelihood function admits no close-form expression and is fairly complex even for small-size hidden Markov models (four or more states), it is hard (if not impossible at all) to show the existence of an isolated maximum, let alone checking the definiteness of the Hessian.

The differentiability, analyticity and other analytic properties of functionals of hidden Markov models similar to the asymptotic likelihood (mainly entropy rate) have recently been studied in [13], [14], [15], [29], [30], [35]. Although very insightful and useful, the results presented in these papers cover only models with discrete state and observation spaces and do not consider the asymptotic behavior of the maximum likelihood estimation method.

In this paper, we study the asymptotic behavior of the recursive maximum likelihood estimation in hidden Markov models with a discrete state-space and continuous observations. We establish a link between the analyticity of the asymptotic log-likelihood on one side, and the point-convergence and convergence rate of the recursive maximum likelihood algorithm, on the other side. More specifically, relying on the principle of analytical continuation, we show under mild conditions that the asymptotic log-likelihood function is analytical in the model parameters if the state transition probabilities and the observation conditional distributions are analytically parameterized. Using this fact and some results from differential geometry (Lojasiewicz inequality), we demonstrate that the recursive maximum likelihood algorithm for hidden Markov models is almost surely point-convergent (i.e., it has a single accumulation point w.p.1). We also derive tight bounds on the almost sure convergence rate. As opposed to all existing results on the asymptotic behavior of maximum likelihood estimation in hidden Markov models, the results of this paper are obtained without assuming that the log-likelihood function has an isolated strong maximum.

The paper is organized as follows. In Section II, the hidden Markov models and the corresponding recursive maximum likelihood algorithms are defined. The main results are presented in Section II, too. Section III provides several practically relevant examples of the main results. Section IV contains the proofs of the main results, while the results of Section III are shown in Section V.

II. Main Results

In order to state the problems of recursive identification and maximum likelihood estimation in hidden Markov models with finite state-spaces and continuous observations, we use the following notation. $N_x > 1$ is an integer, while $\mathcal{X} = \{1, \dots, N_x\}$. $d_y \geq 1$ is also an integer, while $\mathcal{Y}$ is a Borel-measurable set from $\mathbb{R}^{d_y}$. $\{p(x'|x)\}_{x,x' \in \mathcal{X}}$ are non-negative real numbers such that $\sum_{x' \in \mathcal{X}} p(x'|x) = 1$ for each $x \in \mathcal{X}$. $\{Q(\cdot|x)\}_{x \in \mathcal{X}}$ are probability measures on $\mathcal{Y}$. $\{(X_n, Y_n)\}_{n \geq 0}$ is an $\mathcal{X} \times \mathcal{Y}$-valued stochastic process which is defined on a (canonical) probability space $(\Omega, \mathcal{F}, P)$ and satisfies

$$P(Y_{n+1} \in B, X_{n+1} = x | X_0, Y_0, \dots, X_n, Y_n) = Q(B|x) p(x|X_n)$$

w.p.1 for all $x \in \mathcal{X}$, $n \geq 0$, and any Borel measurable set B from $\mathcal{Y}$. On the other side, d_θ is a positive integer, while Θ is an open set from $\mathbb{R}^{d_\theta}$. $\{p_\theta(x'|x)\}_{x,x' \in \mathcal{X}}$ are Borel-measurable functions of $\theta \in \Theta$ such that $p_\theta(x'|x) \geq 0$ and $\sum_{x'' \in \mathcal{X}} p_\theta(x''|x) = 1$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$. $\{q_\theta(y|x)\}_{x \in \mathcal{X}}$ are Borel-measurable functions of $(\theta, y) \in \Theta \times \mathcal{Y}$ such that $q_\theta(y|x) \geq 0$ and $\int_{\mathcal{Y}} q_\theta(y'|x) dy' = 1$ for all $\theta \in \Theta$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. For $\theta \in \Theta$, $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$ is an $\mathcal{X} \times \mathcal{Y}$-valued stochastic process which is defined on a (canonical) probability space $(\Omega, \mathcal{F}, P_\theta)$ and admits

$$P_\theta(Y_{n+1}^\theta \in B, X_{n+1}^\theta = x | X_0^\theta, Y_0^\theta, \dots, X_n^\theta, Y_n^\theta) = \int_B q_\theta(y|x) p_\theta(x|X_n^\theta) dy$$

w.p.1 for each $x \in \mathcal{X}$, $n \geq 0$, and any Borel measurable set B from Y. Finally, $f(\cdot)$ stands for the asymptotic value of the log-likelihood function associated with data $\{Y_n\}_{n \geq 0}$. It is defined by

$$f(\theta) = \lim_{n \to \infty} E\left(\frac{1}{n} \log p_\theta^n(Y_1, \dots, Y_n) \right)$$

for $\theta \in \Theta$, where

$$p_\theta^n(y_1, \dots, y_n) = \sum_{x_0, \dots, x_n \in \mathcal{X}} P_\theta(X_0^\theta = x_0) \prod_{i=1}^n \left(q_\theta(y_k|x_k) p_\theta(x_k|x_{k-1}) \right)$$

for $\theta \in \Theta$, $y_1, \dots, y_n \in \mathcal{Y}$, $n \geq 0$.

In the statistics and engineering literature, $\{(X_n, Y_n)\}_{n \geq 0}$ (as well as $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$) is commonly referred to as a hidden Markov model with a finite state-space and continuous observations, while X_n and Y_n are considered as the (unobservable) state and (observable) output at discrete-time n. On the other hand, the identification of $\{(X_n, Y_n)\}_{n \geq 0}$ is regarded to as the estimation (or approximation) of $\{p(x'|x)\}_{x,x' \in \mathcal{X}}$ and $\{Q(\cdot|x)\}_{x \in \mathcal{X}, y \in \mathcal{Y}}$ given the output sequence $\{Y_n\}_{n \geq 0}$. If the identification is based on the maximum likelihood principle and the parameterized model $\{p_\theta(x'|x)\}_{x,x' \in \mathcal{X}}$, $\{q_\theta(y|x)\}_{x \in \mathcal{X}, y \in \mathcal{Y}}$, the estimation reduces to the maximization of the likelihood function $f(\cdot)$ over Θ. In that context, $\{(X_n^\theta, Y_n^\theta)\}_{n \geq 0}$ is considered as a candidate model of $\{(X_n, Y_n)\}_{n \geq 0}$. For more details on hidden Markov models and their identification see [8, Part II] and references cited therein.

Since the asymptotic mean of $\log p_\theta^n(Y_1, \dots, Y_n)/n$ is rarely available analytically, $f(\cdot)$ is usually maximized by a stochastic gradient algorithm, which itself is a special case of stochastic approximation (for details see [2], [18],

[32] and references cited therein). To define such an algorithm, we introduce some further notation. For $\theta \in \mathbb{R}^{d_\theta}$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, let

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

while $R_\theta(y)$ is an $\mathbb{R}^{N_x \times N_x}$ matrix whose (i,j) entry is $r_\theta(y|i,j)$ (i.e., $R_\theta(y) = [r_\theta(y|i,j)]_{i,j\in\mathcal{X}}$). On the other side, for $\theta \in \mathbb{R}^{d_\theta}$, $u \in [0,\infty)^{N_x} \setminus \{0\}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $y \in \mathcal{Y}$, $1 \leq k \leq d_\theta$, let

$$\begin{aligned}
&\phi_\theta(u,y) = \log(e^T R_\theta(y)u), \\
&F_\theta(u,V,y) = \nabla_\theta \phi_\theta(u,y) + V\nabla_u \phi_\theta(u,y), \\
&G_\theta(u,y) = \frac{R_\theta(y)u}{e^T R_\theta(y)u}, \\
&H_\theta(u,V,y) = \nabla_\theta G_\theta(u,y) + V\nabla_u G_\theta(u,y)
\end{aligned}$$

where $e = [1 \ldots 1]^T \in \mathbb{R}^{N_x}$. With this notation, a stochastic gradient algorithm for maximizing $f(\cdot)$ can be defined as

$$\theta_{n+1} = \theta_n + \alpha_n F_{\theta_n}(U_n, V_n, Y_{n+1}), \tag{1}$$

$$U_{n+1} = G_{\theta_{n+1}}(U_n, Y_{n+1}), \tag{2}$$

$$V_{n+1} = H_{\theta_{n+1}}(U_n, V_n, Y_{n+1}), \quad n \geq 0. \tag{3}$$

In this recursion, $\{\alpha_n\}_{n\geq 0}$ denotes a sequence of positive reals. $\theta_0 \in \mathbb{R}^{d_\theta}$, $U_0 \in \mathbb{R}^{N_x}$ and $V_0 \in \mathbb{R}^{d_\theta \times N_x}$ are random variables which are defined on the probability space $(\Omega, \mathcal{F}, P)$ and independent of $\{Y_n\}_{n\geq 0}$.

In the literature on hidden Markov models and system identification, recursion (1) – (3) is known as the recursive maximum likelihood algorithm, while subrecursions (2) and (3) are referred to as the optimal filter and the optimal filter derivatives, respectively (see [8] for further details). Recursion (1) – (3) usually includes a projection (or truncation) device which prevents estimates $\{\theta_n\}_{n\geq 0}$ from leaving Θ (see [25] for further details). However, in order to avoid unnecessary technical details and to keep the exposition as simple as possible, this aspect of algorithm (1) – (3) is not considered here. Instead, similarly as in [25], our results on the asymptotic behavior of algorithm (1) – (3) (Theorems 2 and 3) are expressed in a local form.

Throughout the paper, unless stated otherwise, the following notation is used. For an integer $d \geq 1$, $\mathcal{P}^d$ denotes the set of d-dimensional probability vectors (i.e., $\mathcal{P}^d = \{u \in [0,\infty)^d : e^T u = 1\}$), while $\mathbb{C}^d$ and $\mathbb{C}^{d\times d}$ are the sets of d-dimensional complex vectors and $d \times d$ complex matrices (respectively). $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$ or $\mathbb{C}^d$, while $d(\cdot,\cdot)$ is the distance induced by this norm. For a real number $\delta \in (0,\infty)$ and a set $A \subseteq \mathbb{C}^d$, $V_\delta(A)$ is the (complex) δ-vicinity of A induced by distance $d(\cdot,\cdot)$, i.e.,

$$V_\delta(A) = \{w \in \mathbb{C}^d : d(w, A) \leq \delta\}.$$

S is the set of stationary points of $f(\cdot)$, i.e.,

$$S = \{\theta \in \Theta : \nabla f(\theta) = 0\}.$$

Algorithm (1) – (3) is analyzed under the following assumptions.

Assumption 1: $\lim_{n\to\infty} \alpha_n = 0$, $\limsup_{n\to\infty} |\alpha_{n+1}^{-1} - \alpha_n^{-1}| < \infty$ and $\sum_{n=0}^{\infty} \alpha_n = \infty$. Moreover, there exists a real number $r \in (1,\infty)$ such that $\sum_{n=0}^{\infty} \alpha_n^2 \gamma_n^{2r} < \infty$.

Assumption 2: $\{X_n\}_{n\geq 0}$ is geometrically ergodic.

Assumption 3: There exists a function $s_\theta(y|x)$ mapping $(\theta,x,y) \in \Theta \times \mathcal{X} \times \mathcal{Y}$ into $[0,\infty)$, and for any compact set $Q \subset \Theta$, there exists a real number $\varepsilon_Q \in (0,1)$ such that

$$\varepsilon_Q s_\theta(y|x') \leq r_\theta(y|x',x) \leq \varepsilon_Q^{-1} s_\theta(y|x')$$

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

Assumption 4: For each $y \in \mathcal{Y}$, $\phi_\theta(u,y)$ and $G_\theta(u,y)$ are real-analytic functions of (θ,u) on entire $\Theta \times \mathcal{P}^{N_x}$. Moreover, $\phi_\theta(u,y)$ and $G_\theta(u,y)$ have (complex-valued) analytical continuations $\hat{\phi}_\eta(w,y)$ and $\hat{G}_\eta(w,y)$ (respectively) with the following properties:

i) $\hat{\phi}_\eta(w,y)$ and $\hat{G}_\eta(w,y)$ map $(\eta,w,y) \in \mathbb{C}^{d_\theta} \times \mathbb{C}^{N_x} \times \mathcal{Y}$ into $\mathbb{C}$ and $\mathbb{C}^{N_x}$ (respectively).

ii) $\hat{\phi}_\theta(u,y) = \phi_\theta(u,y)$ and $\hat{G}_\theta(u,y) = G_\theta(u,y)$ for all $\theta \in \Theta$, $u \in \mathcal{P}^{N_x}$, $y \in \mathcal{Y}$.

iii) For any compact set $Q \subset \Theta$, there exist real numbers $\delta_Q \in (0,1)$, $K_Q \in [1,\infty)$ and a Borel-measurable function $\psi_Q : \mathcal{Y} \to [1,\infty)$ such that $\hat{\phi}_\eta(w,y)$ and $\hat{G}_\eta(w,y)$ are analytical in (η,w) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x})$ for each $y \in \mathcal{Y}$, and such that

$$|\hat{\phi}_\eta(w,y)| \leq \psi_Q(y),$$

$$\|\hat{G}_\eta(w,y)\| \leq K_Q,$$

$$\int \psi_Q^2(y') Q(dy'|x) < \infty$$

for all $\eta \in V_{\delta_Q}(Q)$, $w \in V_{\delta_Q}(\mathcal{P}^{N_x})$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

Assumption 1 corresponds to the properties of step-size sequence $\{\alpha_n\}_{n\geq 0}$ and is commonly used in the asymptotic analysis of stochastic approximation algorithms. It holds if $\alpha_n = 1/n^a$ for $n \geq 1$, where $a \in (3/4, 1]$.

Assumptions 2 and 3 are related to the stability of the model $\{(X_n, Y_n)\}_{n\geq 0}$ and its optimal filter. In this or similar form, they are involved in the analysis of various aspects of optimal filtering and parameter estimation in hidden Markov models (see e.g., [5], [6], [10], [11], [19] – [22], [24], [28], [33], [34], [36]; see also [8, Part II] and references cited therein).

Assumption 4 corresponds to the parametrization of candidate models $\{(X_n^\theta, Y_n^\theta)\}_{n\geq 0}$. Basically, Assumption 4 requires transition probabilities $p_\theta(x'|x)$ and observation conditional densities $q_\theta(y|x)$ to be analytic in θ. It also requires $p_\theta(x'|x)$ and $q_\theta(y|x)$ can be analytically continuable to a complex domain such that the corresponding continuation of the optimal filter transfer function $G_\theta(u,y)$ is analytic and uniformly bounded in (θ,u). Although these requirements are restrictive, they still hold in many practically relevant cases and situations. Several examples are provided in the next section.

In order to state our main results we rely on the following notation. $\{\gamma_n\}_{n\geq 0}$ is a sequence of real numbers defined by $\gamma_0 = 1$ and

$$\gamma_n = 1 + \sum_{i=0}^{n-1} \alpha_i$$

for $n \geq 1$. Event Λ is defined as

$$\Lambda = \left\{ \sup_{n\geq 0} \|\theta_n\| < \infty, \inf_{n\geq 0} d(\theta_n, \partial\Theta) > 0 \right\}.$$

With this notation, our main results on the properties of objective function $f(\cdot)$ and algorithm (1) – (3) can be stated as follows.

Theorem 1 (Analyticity): Let Assumptions 2 – 4 hold. Then, the following is true:

i) $f(\cdot)$ is analytic on entire Θ.

ii) For each $\theta \in \Theta$, there exist real numbers $\delta_\theta \in (0,1)$, $\mu_\theta \in (1,2]$, $M_\theta \in [1,\infty)$ such that

$$|f(\theta') - f(\theta)| \leq M_\theta \|\nabla f(\theta')\|^{\mu_\theta}$$

for all $\theta' \in \Theta$ satisfying $\|\theta - \theta'\| \leq \delta_\theta$.

Theorem 2 (Convergence): Let Assumption 1 – 4 hold. Then, $\hat{\theta} = \lim_{n\to\infty} \theta_n$ exists and satisfies $\nabla f(\hat{\theta}) = 0$ w.p.1 on event Λ.

Theorem 3 (Convergence Rate): Let Assumptions 1 – 4 hold. Then,

$$\|\nabla f(\theta_n)\|^2 = O\big(\gamma_n^{-\hat{p}}\big), \qquad |f(\theta_n) - f(\hat{\theta})| = O\big(\gamma_n^{-\hat{p}}\big), \qquad \|\theta_n - \hat{\theta}\| = O\big(\gamma_n^{-\hat{q}}\big) \tag{4}$$

w.p.1 on Λ, where $\hat{\mu} = \mu_{\hat{\theta}}$ and

$$\hat{r} = \begin{cases} 1/(2-\hat{\mu}), & \text{if } \hat{\mu} < 2 \\ \infty, & \text{otherwise} \end{cases}, \qquad \hat{p} = \hat{\mu}\min\{r, \hat{r}\}, \qquad \hat{q} = \min\{(\hat{p}-1)/2, r-1\}. \tag{5}$$

Proofs of the Theorems 1 – 3 are provided in Section IV.

In the literature on deterministic and stochastic optimization (notice that recursion (1) – (3) belongs to the class of stochastic gradient algorithms), the convergence of gradient search is usually characterized by gradient, objective and estimate convergence, i.e., by the convergence of sequences $\{\nabla f(\theta_n)\}_{n\geq 0}$, $\{f(\theta_n)\}_{n\geq 0}$ and $\{\theta_n\}_{n\geq 0}$ (see e.g., [3], [4], [31], [32] and references cited therein). Similarly, the convergence rate can be described by the rates at which sequences $\{\nabla f(\theta_n)\}_{n\geq 0}$, $\{f(\theta_n)\}_{n\geq 0}$ and $\{\theta_n\}_{n\geq 0}$ converge to the sets of their accumulation points. In the case of algorithm (1) – (3), this kind of information is provided by Theorems 2 and 3. Basically, Theorem 2 claims that recursion (1) – (3) is point-convergent w.p.1 (i.e., the set of accumulation points of $\{\theta_n\}_{n\geq 0}$ is almost surely a singleton), while Theorem 3 provides relatively tight bounds on convergence rate in the terms of Lojasiewicz exponent $\mu_{\hat{\theta}}$ and the convergence rate of step-sizes $\{\alpha_n\}_{n\geq 0}$ (expressed through r and $\{\gamma_n\}_{n\geq 0}$). Theorem 1, on the other side, deals with the properties of the asymptotic log-likelihood $f(\cdot)$ and is a crucial prerequisite for Theorems 2 and 3. Apparently, the results of Theorems 2 and 3 are of local nature: They hold on the event where algorithm (1) – (3) is stable (i.e., where $\{\theta_n\}_{n\geq 0}$ is contained in a compact subset of Θ). Stating asymptotic results in such a form is quite common for stochastic recursive algorithms (see e.g., [18], [25] and references cited therein). Moreover, a global version of Theorems 2 and 3 can be obtained easily by combining them with methods used to verify or ensure stability (e.g., with [7], [9] or [25]).

Various asymptotic properties of maximum likelihood estimation in hidden Markov models have been analyzed thoroughly in a number of papers [1], [5], [6], [10], [11], [19] – [22], [24], [28], [33], [34]; (see also [8, Chapter 12], [12] and references cited therein). Although these results offer a deep insight into the asymptotic behavior of this estimation method, they can hardly be applied to complex hidden Markov models. The reason comes out of the fact that all existing results on the point-convergence and convergence rate of stochastic gradient search (which includes recursive maximum likelihood estimation as a special case) require objective function to have an isolated maximum (or minimum) at which the Hessian is strictly negative definite. Since $f(\cdot)$, the objective function of recursion (1) – (3), is rather complex even when the observation space is finite (i.e., $\mathcal{Y} = \{1, \dots, N_y\}$) and N_x, N_y, the numbers of states and observations, are relatively small (three and above), it is hard (if possible at all) to show the existence of isolated maxima, let alone checking the definiteness of $\nabla^2 f(\cdot)$. Exploiting the analyticity of $f(\cdot)$ and Lojasiewicz inequality, Theorems 2 and 3 overcome these difficulties: They both neither require the existence of an isolated maximum, nor impose any restriction on the definiteness of the Hessian (notice that the Hessian cannot be strictly definite at a non-isolated maximum or minimum). In addition to this, the theorems cover a relatively broad class of hidden Markov models (see the next section). To the best of our knowledge, asymptotic results with similar features do not exist in the literature on hidden Markov models or stochastic optimization.

The differentiability, analyticity and other analytic properties of the entropy rate of hidden Markov models, a functional similar to the asymptotic likelihood, have been studied thoroughly in several papers [13], [14], [15], [29], [30], [35]. The results presented therein cover only models with discrete state and observation spaces and do not pay any attention to maximum likelihood estimation. Motivated by the problem of the point-convergence and convergence rate of recursive maximum likelihood estimators for hidden Markov models, we extend these results in Theorem 1 to models with continuous observations and their likelihood functions. The approach we use to demonstrate the analyticity of the asymptotic likelihood is based on the principle of analytical continuation and is similar to the methodology formulated in [13].

III. Examples

In this section, we consider several practically relevant examples of the results presented in Section II. Analyzing these examples, we also provide a direction how the assumptions adopted in Section II can be verified in practice.

A. *Finite Observation Space*

Hidden Markov models with finite state and observation spaces are studied in this subsection. For these models, we show that the conclusion of Theorems 1 – 3 hold whenever the parameterization of candidate models is analytic.

Let $N_y > 2$ be an integer, while $\mathcal{Y} = \{1, \dots, N_y\}$. Then, the following results hold.

Proposition 1: Assumptions 3 and 4 are true if the following conditions are satisfied:

i) For each $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, $r_\theta(y|x', x)$ is analytical in θ on entire Θ.

ii) $r_\theta(y|x', x) > 0$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

Corollary 1: Let Assumptions 1, 2 and the conditions of Proposition 1 hold. Then, the conclusions of Theorems 1 – 3 are true.

The proof is provided in Section V.

Remark: The conditions of Proposition 1 correspond to the way the candidate models are parameterized. They hold for the natural[1], exponential[2] and trigonometric[3] parameterizations.

B. Compactly Supported Observations

In this subsection, we consider hidden Markov models with a finite number of states and compactly supported observations. More specifically, we assume that $\mathcal{Y}$ is a compact set from $\mathbb{R}^{d_y}$. For such models, the following results can be shown.

Proposition 2: Assumptions 3 and 4 are true if the following conditions are satisfied:

i) For each $x, x' \in \mathcal{X}$, $r_\theta(y|x', x)$ is analytical in (θ, y) on entire $\Theta \times \mathcal{Y}$.

ii) $r_\theta(y|x', x) > 0$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

Corollary 2: Let Assumptions 1, 2 and the conditions of Proposition 2 hold. Then, the conclusions of Theorems 1 – 3 are true.

The proof is provided in Section V.

Remark: The conditions of Proposition 2 are fulfilled if the natural, exponential or trigonometric parameterization (see the previous subsection) is applied to the state transition probabilities $\{p_\theta(x'|x)\}_{x,x' \in \mathcal{X}}$, and if the observation likelihoods $\{q_\theta(\cdot|x)\}_{x \in \mathcal{X}}$ are analytic jointly in θ and y. The later holds when $\{q_\theta(\cdot|x)\}_{x \in \mathcal{X}}$ are compactly truncated mixtures of beta, exponential, gamma, logistic, normal, log-normal, Pareto, uniform, Weibull distributions, and when each of these mixtures is indexed by its weights and by the 'natural' parameters of its ingredient distributions.

[1] The natural parameterization can be defined as follows: $\theta = [\alpha_{1,1} \cdots \alpha_{N_x,N_x} \beta_{1,1} \cdots \beta_{N_x,N_y}]^T$ and $p_\theta(x'|x) = \alpha_{x,x'}$, $q_\theta(y|x) = \beta_{x,y}$ for $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, while Θ is the set of vectors $[\alpha_{1,1} \cdots \alpha_{N_x,N_x} \ \beta_{1,1} \cdots \beta_{N_x,N_y}]^T \in (0,1)^{N_x(N_x+N_y)}$ satisfying $\sum_{l=1}^{N_x} \alpha_{x,l} = \sum_{l=1}^{N_y} \beta_{x,l} = 1$ for each $x \in \mathcal{X}$.

[2] In the case of the exponential parameterization, we have $\theta = [\alpha_{1,1} \cdots \alpha_{N_x,N_x} \ \beta_{1,1} \cdots \beta_{N_x,N_y}]^T$, and

$$p_\theta(x'|x) = \frac{\exp(\alpha_{x,x'})}{\sum_{l=1}^{N_x} \exp(\alpha_{x,l})}, \quad q_\theta(y|x) = \frac{\exp(\beta_{x,y})}{\sum_{l=1}^{N_y} \exp(\beta_{x,l})}$$

for $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, while $\Theta = \mathbb{R}^{N_x(N_x+N_y)}$.

[3] The trigonometric parameterization is defined as $\theta = [\alpha_{1,1} \cdots \alpha_{N_x,N_x} \ \beta_{1,1} \cdots \beta_{N_x,N_y}]^T$ and

$$p_\theta(1|x) = \cos^2 \alpha_{x,1}, \quad p_\theta(x'|x) = \cos^2 \alpha_{x,x'} \prod_{l=1}^{x'-1} \sin^2 \alpha_{x,l}, \quad p_\theta(N_x|x) = \prod_{l=1}^{N_x} \sin^2 \alpha_{x,l},$$

$$q_\theta(1|x) = \cos^2 \beta_{x,1}, \quad q_\theta(y|x) = \cos^2 \beta_{x,y} \prod_{l=1}^{y-1} \sin^2 \beta_{x,l}, \quad q_\theta(N_y|x) = \prod_{l=1}^{N_y} \sin^2 \beta_{x,l}$$

for $x \in \mathcal{X}$, $x' \in \mathcal{X} \setminus \{1, N_x\}$, $y \in \mathcal{Y} \setminus \{1, N_y\}$, while $\Theta = (0, \pi/2)^{N_x(N_x+N_y)}$.

C. Mixture of Observation Likelihoods

In this subsection, we consider the case when the observation likelihoods $\{q_\theta(\cdot|x)\}_{x\in\mathcal{X}}$ are mixtures of known probability density functions. More specifically, let $d_\alpha \geq 1$, $N_\beta > 1$ be integers, while $\mathcal{A} \subseteq \mathbb{R}^{d_\alpha}$ is an open set and

$$\mathcal{B} = \left\{ [\beta_{1,1} \cdots \beta_{N_x,N_\beta}]^T \in (0,1)^{N_x N_\beta} : \sum_{i=1}^{N_\beta} \beta_{x,k} = 1 \text{ for each } x \in \mathcal{X} \right\}.$$

We assume that the state transition probabilities and the observation likelihoods are parameterized by vectors $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$ (respectively), i.e., $p_\theta(x'|x) = p_\alpha(x'|x)$, $q_\theta(y|x) = q_\beta(y|x)$ for $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$, $\theta = [\alpha^T \, \beta^T]^T$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. We also assume

$$q_\beta(y|x) = \sum_{k=1}^{N_\beta} \beta_{x,k} f_k(y|x),$$

where $\beta = [\beta_{1,1} \cdots \beta_{N_x,N_\beta}]^T \in \mathcal{B}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, while $\{f_k(\cdot|x)\}_{x\in\mathcal{X}, 1\leq k\leq N_\beta}$ are known probability density functions.

For the models specified in this subsection, the following results hold.

Proposition 3: Assumptions 3 and 4 are true if the following conditions are satisfied:

i) For each $x, x' \in \mathcal{X}$, $p_\alpha(x'|x)$ is analytical in α on entire $\mathcal{A}$.

ii) $p_\alpha(x'|x) > 0$ for all $\alpha \in \mathcal{A}$, $x, x' \in \mathcal{X}$.

iii) $\psi(y) > 0$ and $\int \log^2 \psi(y') Q(dy'|x) < \infty$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, where $\psi(y) = \sum_{x\in\mathcal{X}} \sum_{k=1}^{N_\beta} f_k(y|x)$.

Corollary 3: Let Assumptions 1, 2 and the conditions of Proposition 3 hold. Then, the conclusions of Theorems 1 – 3 are true.

The proof is provided in Section V.

D. Gaussian Observations

This subsection is devoted to hidden Markov models with a finite number of states and with Gaussian observations. More specifically, let d_α and $\mathcal{A}$ have the same meaning as in the previous section, while $\mathcal{Y} = \mathbb{R}$ and

$$\mathcal{B} = \left\{ [\lambda_1 \cdots \lambda_{N_x} \, \mu_1 \cdots \mu_{N_x}]^T \in (0,\infty)^{N_x} \times \mathbb{R}^{N_x} : \lambda_x \neq \lambda_x \text{ for } x \neq x', x, x' \in \mathcal{X} \right\}. \tag{6}$$

Similarly as in the previous subsection, we assume that the state transition probabilities and the observation likelihoods are indexed by vectors $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$ (respectively). We also assume

$$q_\beta(y|x) = \sqrt{\lambda_x/\pi} \exp(-\lambda_x (y - \mu_x)^2),$$

where $\beta = [\lambda_1 \cdots \lambda_{N_x} \, \mu_1 \cdots \mu_{N_x}]^T \in \mathcal{B}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

For the models described in this subsection, the following results can be shown.

Proposition 4: Assumptions 3 and 4 are true if the following conditions are satisfied:

i) For each $x, x' \in \mathcal{X}$, $p_\alpha(x'|x)$ is analytical in α on entire $\mathcal{A}$.

ii) $p_\alpha(x'|x) > 0$ for all $\alpha \in \mathcal{A}$, $x, x' \in \mathcal{X}$.

iii) $\int y^4 Q(dy|x) < \infty$ for all $x \in \mathcal{X}$.

Corollary 4: Let Assumptions 1, 2 and the conditions of Proposition 4 hold. Then, the conclusions of Theorems 1 – 3 are true.

The proof is provided in Section V.

Remark: Unfortunately, Proposition 4 and Corollary 4 cannot be extended to the case $\mathcal{B} = (0,\infty)^{N_x} \times \mathbb{R}^{N_x}$, since the models specified in the subsection do not satisfy Assumption 4 without the condition $\lambda_x \neq \lambda_{x'}$ for $x \neq x'$ (which appears in (6)).[4] However, this condition is not so restrictive in practice as $\mathcal{B}$ is dense in $(0,\infty)^{N_x} \times \mathbb{R}^{N_x}$ and provides an arbitrarily close approximation to $(0,\infty)^{N_x} \times \mathbb{R}^{N_x}$.

IV. Proof of Main Results

A. Optimal Filter and Its Properties

The stability properties (forgetting and ergodicity) of the optimal filter (2), its derivatives (3) and its analytical continuation (to be defined in the next paragraph) are studied in this subsection. The analysis mainly follows the ideas and results of [21], [22] and [23]. The results presented in the subsection are an essential prerequisite for the analysis carried out in Subsections IV-B and IV-C.

Throughout this subsection, we rely on the following notation. $\mathcal{Q}^{N_x}$ denotes the set

$$\mathcal{Q}^{N_x} = \{u \in [0,\infty)^{N_x} : e^T u \geq 1/2\},$$

where $e = [1 \cdots 1]^T \in R^{N_x}$ ($\mathcal{Q}^{N_x}$ can be any compact set from $[0,\infty)^{N_x}$ satisfying $0 \notin \mathcal{Q}^{N_x}$, int$\mathcal{P}^{N_x} \subset \mathcal{Q}^{N_x}$, but the above one is selected for analytical convenience). For $n \geq m \geq 0$ and a sequence $\boldsymbol{y} = \{y_n\}_{n\geq 0}$ from $\mathcal{Y}$, $y_{m:n}$ denotes finite subsequence $(y_m,\dots,y_n)$. For $u \in [0,\infty)^{N_x} \setminus \{0\}$, $w \in \mathbb{C}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq m \geq 0$ and sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{\eta} = \{\eta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from Θ, $\mathbb{C}^{d_\theta}$, $\mathcal{Y}$ (respectively), let $G^{m:m}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u) = u$, $\hat{G}^{m:m}_{\boldsymbol{\eta},\boldsymbol{y}}(w) = w$, $H^{m:m}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u,V) = V$ and

$$G^{m:n+1}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u) = G_{\vartheta_{n+1}}(G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u), y_{n+1}),$$
$$\hat{G}^{m:n+1}_{\boldsymbol{\eta},\boldsymbol{y}}(w) = \hat{G}_{\eta_{n+1}}(\hat{G}^{m:n}_{\boldsymbol{\eta},\boldsymbol{y}}(w), y_{n+1}),$$
$$H^{m:n+1}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u,V) = H_{\vartheta_{n+1}}(G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u), H^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u,V), y_{n+1})$$

[4]Let $h_{\alpha,y,u}(\beta) = e^T R_\theta(y) u$ for $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$, $\theta = [\alpha^T \beta^T]^T$, $y \in \mathcal{Y}$, $u \in \mathcal{P}^{N_x}$. Obviously, for any $\alpha \in \mathcal{A}$, $y \in \mathcal{Y}$, $u \in \mathcal{P}^{N_x}$, $h_{\alpha,y,u}(\cdot)$ has a unique (complex-valued) analytical continuation, which can be defined as

$$\hat{h}_{\alpha,y,u}(b) = \sum_{x,x' \in \mathcal{X}} \sqrt{l_{x'}/\pi} \exp(-l_{x'}(y - m_{x'})^2) p_\alpha(x'|x) u_x$$

where $b = [l_1 \cdots l_{N_x}\ m_1 \cdots m_{N_x}]^T \in \mathbb{C}^{2N_x}$. Let $\beta = [\lambda_1 \cdots \lambda_{N_x}\ \mu_1 \cdots \mu_{N_x}]^T \in (0,\infty)^{N_x} \times \mathbb{R}^{N_x}$ be any vector satisfying $\lambda_x = \lambda_{x'}$ for some $x \neq x'$, $x, x' \in \mathcal{X}$. Then, it is not hard to deduce that there exist $\alpha \in \mathcal{A}$, $y \in \mathcal{Y}$, $u \in \mathcal{P}^{N_x}$ (depending on β) such that $\hat{h}_{\alpha,y,u}(\cdot)$ has a zero in any (complex) vicinity of β. Since the zeros of the analytical continuation of $e^T R_\theta(y) u$ would be the poles of the analytical continuation of $G_\theta(u,y)$, it is not possible to continue $G_\theta(u,y)$ analytically in any vicinity of point (θ, u), where $\theta = [\alpha^T \beta^T]^T$. Hence, Proposition 4 and Corollary 4 cannot be extended to the case $\mathcal{B} = (0,\infty)^{N_x} \times \mathbb{R}^{N_x}$.

($G_\theta(u,y)$, $\hat{G}_\eta(w,y)$, $H_\theta(u,V,y)$ are defined in Section II). If $\boldsymbol{\vartheta}=\{\theta\}_{n\geq 0}$ (i.e., $\vartheta_n=\theta$), we also use notation $G_{\theta,\boldsymbol{y}}^{m:n}(u)=G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)$, $H_{\theta,\boldsymbol{y}}^{m:n}(u,V)=H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u,V)$, as well as $G_\theta^{0:n}(u,y_{1:n})=G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{0:n}(u)$, $H_\theta^{0:n}(u,V,y_{1:n})=H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{0:n}(u,V)$. Similarly, if $\boldsymbol{\eta}=\{\eta\}_{n\geq 0}$ (i.e., $\eta_n=\eta$), we rely on notation $\hat{G}_{\eta,\boldsymbol{y}}^{m:n}(w)=\hat{G}_{\boldsymbol{\eta},\boldsymbol{y}}^{m:n}(w)$ and $\hat{G}_\eta^{0:n}(w,y_{1:n})=\hat{G}_{\boldsymbol{\eta},\boldsymbol{y}}^{0:n}(w)$. Then, it straightforward to verify

$$\begin{aligned}
&G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)=G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{k:n}(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:k}(u)),\\
&\hat{G}_{\boldsymbol{\eta},\boldsymbol{y}}^{m:n}(w)=\hat{G}_{\boldsymbol{\eta},\boldsymbol{y}}^{k:n}(\hat{G}_{\boldsymbol{\eta},\boldsymbol{y}}^{m:k}(w)),\\
&H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u,V)=H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{k:n}(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:k}(u),H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:k}(u,V))
\end{aligned}$$

for each $u\in[0,\infty)^{N_x}\setminus\{0\}$, $w\in\mathbb{C}^{N_x}$, $V\in\mathbb{R}^{d_\theta\times N_x}$, $0\leq m\leq k\leq n$ and any sequences $\boldsymbol{\vartheta}=\{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{\eta}=\{\eta_n\}_{n\geq 0}$, $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ from Θ, $\mathbb{C}^{d_\theta}$, $\mathcal{Y}$ (respectively). Moreover, it can be demonstrated easily

$$\hat{G}_\eta^{0:n}(w,y_{1:n})=\hat{G}_\eta^{0:k}(\hat{G}_\eta^{0:n-k}(w,y_{1:n-k}),y_{n-k+1:n}) \tag{7}$$

for all $\eta\in\mathbb{C}^{d_\theta}$, $w\in\mathbb{C}^{N_x}$, $0\leq k\leq n$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ from $\mathcal{Y}$. It is also easy to show

$$\begin{aligned}
H_\theta^{0:n}(u,V,y_{1:n})=&\,V\,(\nabla_u G_\theta^{0:n})(u,y_{1:n})+(\nabla_\theta G_\theta^{0:n})(u,y_{1:n}),\\
F_\theta\left(G_\theta^{0:n}(u,y_{1:n}),H_\theta^{0:n}(u,V,y_{1:n}),y_{n+1}\right)=&\,V\,(\nabla_u G_\theta^{0:n})(u,y_{1:n})\,(\nabla_u\phi_\theta)(G_\theta^{0:n}(u,y_{1:n}),y_{n+1})\\
&+\nabla_\theta\big(\phi_\theta(G_\theta^{0:n}(u,y_{1:n}),y_{n+1})\big)
\end{aligned} \tag{8}$$

for each $\theta\in\Theta$, $u\in[0,\infty)^{N_x}\setminus\{0\}$, $V\in\mathbb{R}^{d_\theta\times N_x}$, $n\geq 0$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ from $\mathcal{Y}$ ($\phi_\theta(u,y)$, $F_\theta(u,V,y)$ are Section II; $(\nabla_u G_\theta^{0:n})(u,y)$, $(\nabla_\theta G_\theta^{0:n})(u,y)$ denote the Jacobians of $G_\theta^{0:n}(u,y)$ with respect to u, θ, while $(\nabla_u\phi_\theta)(u,y)$ stands for the gradient of $\phi_\theta(u,y)$ with respect to u).

Besides the previously introduced notation, the following notation is also used in this section. For $u\in[0,\infty)^{N_x}\setminus\{0\}$, $n>m\geq 0$ and sequences $\boldsymbol{\vartheta}=\{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ from Θ, $\mathcal{Y}$ (respectively), let $A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{n:n}(u)=I\in\mathbb{R}^{N_x\times N_x}$ (I denotes a unit matrix) and

$$A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)=(\nabla_u G_{\vartheta_{m+1}})(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:m}(u),y_{m+1})\cdots(\nabla_u G_{\vartheta_n})(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n-1}(u),y_n).$$

Then, it is easy to demonstrate

$$H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u,V)=V\,A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)+\sum_{i=m}^{n-1}(\nabla_\theta G_{\vartheta_{i+1}})(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i}(u),y_{i+1})\,A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{i+1:n}(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i+1}(u)) \tag{9}$$

for each $u\in[0,\infty)^{N_x}\setminus\{0\}$, $V\in\mathbb{R}^{d_\theta\times N_x}$, $n\geq m\geq 0$ and any sequences $\boldsymbol{\vartheta}=\{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ from Θ, $\mathcal{Y}$ (respectively).

In this subsection, we also rely on the following notation. $\mathcal{S}_z$ and $\mathcal{S}_\zeta$ denote sets $\mathcal{S}_z=\mathcal{X}\times\mathcal{Y}\times\mathcal{P}^{N_x}\times\mathbb{R}^{d_\theta\times N_x}$ and $\mathcal{S}_\zeta=\mathcal{X}\times\mathcal{Y}\times\mathcal{P}^{N_x}$. For $\theta\in\Theta$, $P_\theta(\cdot,\cdot)$ and $\tilde{P}_\theta(\cdot,\cdot)$ are the transition kernels of Markov chains

$$\{X_{n+1},Y_{n+1},G_\theta^{0:n}(u,Y_{1:n}),H_\theta^{0:n}(u,V,Y_{1:n})\}_{n\geq 0}\quad\text{and}\quad\{X_n,Y_n,G_\theta^{0:n}(u,Y_{1:n}),H_\theta^{0:n}(u,V,Y_{1:n})\}_{n\geq 0}$$

(respectively), while $\Pi_\theta(\cdot,\cdot)$ and $\tilde{\Pi}_\theta(\cdot,\cdot)$ are the transition kernels of Markov chains

$$\{X_{n+1},Y_{n+1},G_\theta^{0:n}(u,Y_{1:n})\}_{n\geq 0}\quad\text{and}\quad\{X_n,Y_n,G_\theta^{0:n}(u,Y_{1:n})\}_{n\geq 0}$$

(notice that $P_\theta(\cdot,\cdot)$, $\tilde{P}_\theta(\cdot,\cdot)$, $\Pi_\theta(\cdot,\cdot)$, $\tilde{\Pi}_\theta(\cdot,\cdot)$ do not depend on u, V). For $\theta \in \Theta$, $z = (x,y,u,V) \in \mathcal{S}_z$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, let

$$\tilde{F}_\theta(u,V,x) = E(F_\theta(u,V,Y_2)|X_1 = x),$$
$$\tilde{\phi}_\theta(u,x) = E(\phi_\theta(u,Y_2)|X_1 = x)$$

while

$$F(\theta,z) = F_\theta(u,V,y), \quad \tilde{F}(\theta,z) = \tilde{F}_\theta(u,V,y), \quad \phi(\theta,\zeta) = \phi_\theta(u,y), \quad \tilde{\phi}(\theta,\zeta) = \tilde{\phi}_\theta(u,y).$$

Then, it is straightforward to verify

$$\begin{aligned}
(P^n F)(\theta,z) =& E\left(F_\theta(G_\theta^{0:n}(u,Y_{1:n}), H_\theta^{0:n}(u,V,Y_{1:n}), Y_{n+1})|X_1 = x, Y_1 = y\right) \\
=& E\left(\tilde{F}_\theta(G_\theta^{0:n}(u,Y_{1:n}), H_\theta^{0:n}(u,V,Y_{1:n}), X_n)|X_1 = x, Y_1 = y\right) \\
=& (\tilde{P}^{n-1}\tilde{F})\Big(\theta, \big(\, x, y, G_\theta(u,y), H_\theta(u,V,y)\,\big)\Big),
\end{aligned} \tag{10}$$

$$\begin{aligned}
(\Pi^n \phi)(\theta,\zeta) =& E\left(\phi_\theta(G_\theta^{0:n}(u,Y_{1:n}), Y_{n+1})|X_1 = x, Y_1 = y\right) \\
=& E\left(\tilde{\phi}_\theta(G_\theta^{0:n}(u,Y_{1:n}), X_n)|X_1 = x, Y_1 = y\right) \\
=& (\tilde{\Pi}^{n-1}\tilde{\phi})\Big(\theta, \big(\, x, y, G_\theta(u,y)\,\big)\Big)
\end{aligned} \tag{11}$$

for all $\theta \in \Theta$, $z = (x,y,u,V) \in \mathcal{S}_z$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $n > 1$. It can also be concluded

$$\begin{aligned}
& E\left(\frac{\log p_\theta^{n+1}(Y_1,\dots,Y_{n+1})}{n+1} \,\middle|\, X_1 = x, Y_1 = y \right) \\
&= E\left(\frac{1}{n+1} \sum_{i=0}^{n} \phi_\theta(G_\theta^{0:i}(u_\theta, Y_{1:i}), Y_{i+1}) \,\middle|\, X_1 = x, Y_1 = y \right) \\
&= \frac{1}{n+1} \sum_{i=1}^{n} (\tilde{\Pi}^{i-1}\tilde{\phi})\Big(\theta, \big(\, x, y, G_\theta(u_\theta, y)\,\big)\Big) + \frac{\phi_\theta(u_\theta, Y_1)}{n+1}
\end{aligned} \tag{12}$$

for each $\theta \in \Theta$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $n > 1$, where $u_\theta = [P(X_1^\theta = 1) \cdots P(X_1^\theta = N_x)]^T$.

Lemma 1: Suppose that Assumption 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, there exist real numbers $\delta_{1,Q} \in (0,1)$, $C_{1,Q} \in [1,\infty)$ such that

$$|\tilde{\phi}_\theta(u,x)| \leq C_{1,Q}, \tag{13}$$

$$\|F_\theta(u,V,y)\| \leq C_{1,Q}\psi_Q(y)(1+\|V\|), \tag{14}$$

$$\|\tilde{F}_\theta(u,V,x)\| \leq C_{1,Q}(1+\|V\|), \tag{15}$$

$$|\tilde{\phi}_{\theta'}(u',x) - \tilde{\phi}_{\theta''}(u'',x)| \leq C_{1,Q}(\|\theta'-\theta''\| + \|u'-u''\|), \tag{16}$$

$$|\hat{\phi}_{\eta'}(w',y) - \hat{\phi}_{\eta''}(w'',y)| \leq C_{1,Q}\psi_Q(y)(\|\eta'-\eta''\| + \|w'-w''\|), \tag{17}$$

$$\begin{aligned}&\|F_{\theta'}(u',V',y) - F_{\theta''}(u'',V'',y)\| \\ &\quad \leq C_{1,Q}\psi_Q(y)(1+\|V'\|+\|V''\|)(\|\theta'-\theta''\| + \|u'-u''\| + \|V'-V''\|),\end{aligned} \tag{18}$$

$$\begin{aligned}&\|\tilde{F}_{\theta'}(u',V',x) - \tilde{F}_{\theta''}(u'',V'',x)\| \\ &\quad \leq C_{1,Q}(1+\|V'\|+\|V''\|)(\|\theta'-\theta''\| + \|u'-u''\| + \|V'-V''\|)\end{aligned} \tag{19}$$

for all $\theta,\theta',\theta'' \in Q$, $\eta',\eta'' \in V_{\delta_{1,Q}}(Q)$, $u,u',u'' \in \mathcal{P}^{N_x}$, $w',w'' \in V_{\delta_{1,Q}}(\mathcal{P}^{N_x})$, $V,V',V'' \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ ($\psi_Q(\cdot)$ is specified in Assumption 4).

Proof: Let $\delta_{1,Q} = \delta_Q/2$ (δ_Q is defined in Assumption 4). Then, Cauchy inequality for analytic functions (see e.g., [37, Proposition 2.1.3]) and Assumption 4 imply that there exists a real number $\tilde{C}_{1,Q} \in [1,\infty)$ such that

$$\max\{\|\nabla_{(\eta,w)}\hat{\phi}_\eta(w,y)\|, \|\nabla^2_{(\eta,w)}\hat{\phi}_\eta(w,y)\|\} \leq \tilde{C}_{1,Q}\psi_Q(y)$$

for all $\eta \in V_{\delta_{1,Q}}(Q)$, $w \in V_{\delta_{1,Q}}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$ ($\nabla_{(\eta,w)}$, $\nabla^2_{(\eta,w)}$ denote the gradient and Hessian with respect to (η,w)). Consequently, there exists another real number $\tilde{C}_{2,Q} \in [1,\infty)$ such that

$$\begin{aligned}&\max\{\|\hat{\phi}_{\eta'}(w',y) - \hat{\phi}_{\eta''}(w'',y)\|, \|\nabla_w\hat{\phi}_{\eta'}(w',y) - \nabla_w\hat{\phi}_{\eta''}(w'',y)\|\} \\ &\leq \tilde{C}_{2,Q}\psi_Q(y)(\|\eta'-\eta''\| + \|w'-w''\|)\end{aligned}$$

for any $\eta',\eta'' \in V_{\delta_{1,Q}}(Q)$, $w',w'' \in V_{\delta_{1,Q}}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$. Therefore,

$$\begin{aligned}\|F_\theta(u,V,y)\| \leq& \|\nabla_\theta\phi_\theta(u,y)\| + \|\nabla_u\phi_\theta(u,y)\|\|V\| \\ \leq& \tilde{C}_{1,Q}\psi_Q(y)(1+\|V\|), \\ \|F_{\theta'}(u',V',y) - F_{\theta''}(u'',V'',y)\| \leq& \|\nabla_\theta\phi_{\theta'}(u',y) - \nabla_\theta\phi_{\theta''}(u'',y)\| + \|\nabla_u\phi_{\theta'}(u',y) - \nabla_u\phi_{\theta''}(u'',y)\|\|V'\| \\ &+ \|\nabla_u\phi_{\theta''}(u'',y)\|\|V'-V''\| \\ \leq& \tilde{C}_{2,Q}\psi_Q(y)(1+\|V'\|+\|V''\|)(\|\theta'-\theta''\| + \|u'-u''\|) \\ &+ \tilde{C}_{1,Q}\psi_Q(y)\|V'-V''\|\end{aligned}$$

for each $\theta, \theta', \theta'' \in Q$, $u, u', u'' \in \mathcal{P}^{N_x}$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$. We also have

$$\|\tilde{F}_\theta(u, V, x)\| \leq \tilde{C}_{1,Q}(1 + \|V\|) \int \psi_Q(y) Q(dy|x)$$

$$\|\tilde{F}_{\theta'}(u', V', x) - \tilde{F}_{\theta''}(u'', V'', x)\|$$

$$\leq (\tilde{C}_{1,Q} + \tilde{C}_{2,Q})(1 + \|V'\| + \|V''\|)(\|\theta' - \theta''\| + \|u' - u''\| + \|V' - V''\|) \int \psi_Q(y) Q(dy|x)$$

for all $\theta, \theta', \theta'' \in Q$, $u, u', u'' \in \mathcal{P}^{N_x}$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$. Then, it can be deduced that there exists a real number $C_{1,Q} \in [1, \infty)$ such that (13) – (19) hold for each $\theta, \theta', \theta'' \in Q$, $\eta', \eta'' \in V_{\delta_{1,Q}}(Q)$, $u, u', u'' \in \mathcal{P}^{N_x}$, $w', w'' \in V_{\delta_{1,Q}}(\mathcal{P}^{N_x})$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. ∎

Lemma 2: Suppose that Assumption 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, there exist real numbers $\delta_{2,Q} \in (0, 1)$, $C_{2,Q} \in [1, \infty)$ such that

$$\|\nabla_\eta \hat{G}_\eta(w, y)\| \leq C_{2,Q}, \tag{20}$$

$$\|H_\theta(u, V, y)\| \leq C_{2,Q}(1 + \|V\|), \tag{21}$$

$$\max\{\|\hat{G}_{\eta'}(w', y) - \hat{G}_{\eta''}(w'', y)\|, \|\nabla_w \hat{G}_{\eta'}(w', y) - \nabla_w \hat{G}_{\eta''}(w'', y)\|\}$$

$$\leq C_{2,Q}(\|\eta' - \eta''\| + \|w' - w''\|), \tag{22}$$

$$\|H_{\theta'}(u, V, y) - H_{\theta''}(u, V, y)\| \leq C_{2,Q}(1 + \|V\|)\|\theta' - \theta''\| \tag{23}$$

for all $\theta, \theta', \theta'' \in Q$, $\eta, \eta', \eta'' \in V_{\delta_{2,Q}}(Q)$, $u \in \mathcal{P}^{N_x}$, $w, w', w'' \in V_{\delta_{2,Q}}(\mathcal{P}^{N_x})$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $y \in \mathcal{Y}$.

Proof: Let $\delta_{2,Q} = \min\{\delta_Q/2, \delta_{1,Q}\}$. Owing to Cauchy inequality for analytic functions and Assumption 4, there exists a real number $\tilde{C}_{1,Q} \in [1, \infty)$ such that

$$\max\{\|\nabla_{(\eta,w)} \hat{G}_\eta^k(w, y)\|, \|\nabla^2_{(\eta,w)} \hat{G}_\eta^k(w, y)\|\} \leq \tilde{C}_{1,Q}$$

for any $\eta \in V_{\delta_{2,Q}}(Q)$, $w \in V_{\delta_{2,Q}}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$ ($\hat{G}_\eta^k(w, y)$ stands for the k-th component of $\hat{G}_\eta(w, y)$). Consequently, there exists another real number $\tilde{C}_{2,Q} \in [1, \infty)$ such that

$$\max\{\|\hat{G}_{\eta'}(w', y) - \hat{G}_{\eta''}(w'', y)\|, \|\nabla_\eta \hat{G}_{\eta'}(w', y) - \nabla_\eta \hat{G}_{\eta''}(w'', y)\|, \|\nabla_w \hat{G}_{\eta'}(w', y) - \nabla_w \hat{G}_{\eta''}(w'', y)\|\}$$

$$\leq \tilde{C}_{2,Q}(\|\eta' - \eta''\| + \|w' - w''\|)$$

for all $\eta', \eta'' \in V_{\delta_{2,Q}}(Q)$, $w', w'' \in V_{\delta_{2,Q}}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$. Therefore,

$$\begin{aligned} \|H_\theta(u, V, y)\| \leq& \|\nabla_\theta G_\theta(u, y)\| + \|\nabla_u G_\theta(u, y)\| \|V\| \\ \leq& \tilde{C}_{1,Q} N_x (1 + \|V\|), \end{aligned}$$

$$\begin{aligned} \|H_{\theta'}(u, V, y) - H_{\theta''}(u, V, y)\| \leq& \|\nabla_\theta G_{\theta'}(u, y) - \nabla_\theta G_{\theta''}(u, y)\| + \|\nabla_u G_{\theta'}(u, y) - \nabla_u G_{\theta''}(u, y)\| \|V\| \\ \leq& \tilde{C}_{2,Q}(1 + \|V\|)\|\theta' - \theta''\| \end{aligned}$$

for each $\theta, \theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$. Then, it is clear that there exists a real number $C_{2,Q} \in [1, \infty)$ such that (20) – (23) hold for all $\theta, \theta', \theta'' \in Q$, $\eta, \eta', \eta'' \in V_{\delta_{2,Q}}(Q)$, $u \in \mathcal{P}^{N_x}$, $w, w', w'' \in V_{\delta_{2,Q}}(\mathcal{P}^{N_x})$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $y \in \mathcal{Y}$. ∎

Lemma 3: Suppose that Assumptions 3 and 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, the following is true:

i) There exist real numbers $\varepsilon_{1,Q} \in (0,1)$, $C_{3,Q} \in [1,\infty)$ such that

$$\|A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)\| \leq C_{3,Q}\varepsilon_{1,Q}^{n-m}, \tag{24}$$

$$\|A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u') - A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u'')\| \leq C_{3,Q}\varepsilon_{1,Q}^{n-m}\|u'-u''\|, \tag{25}$$

$$\|G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(w') - G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(w'')\| \leq C_{3,Q}\varepsilon_{1,Q}^{n-m}\|w'-w''\| \tag{26}$$

for all $u, u', u'' \in \mathcal{P}^{N_x}$, $w', w'' \in \mathcal{Q}^{N_x}$, $n \geq m \geq 0$ and any sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from Q, $\mathcal{Y}$ (respectively).

ii) There exist real numbers $\varepsilon_{2,Q} \in (0,1)$, $C_{4,Q} \in [1,\infty)$ such that

$$\|H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u,V)\| \leq C_{4,Q}(1+\|V\|) \tag{27}$$

$$\|H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u',V') - H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u'',V'')\| \leq C_{4,Q}\varepsilon_{2,Q}^{n-m}\big(\|u'-u''\|(1+\|V'\|+\|V''\|) + \|V'-V''\|\big) \tag{28}$$

for all $u, u', u'' \in \mathcal{P}^{N_x}$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq m \geq 0$ and any sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from Q, $\mathcal{Y}$ (respectively).

Proof: Using [36, Theorem 3.1, Lemmas 6.6, 6.7] (with a few straightforward modifications), it can be deduced from Assumption 3 that there exist real numbers $\varepsilon_{1,Q} \in (0,1)$, $C_{3,Q} \in [1,\infty)$ such that (24), (25) and

$$\|G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(w') - G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(w'')\| \leq 2^{-1}(N_x+1)^{-1}C_{3,Q}\varepsilon_{1,Q}^{n-m}\left\|\frac{w'}{e^T w'} - \frac{w''}{e^T w''}\right\|$$

hold for all $u, u', u'' \in \mathcal{P}^{N_x}$, $w', w'' \in [0,\infty)^{N_x} \setminus \{0\}$, $n \geq m \geq 0$ and any sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from Q, $\mathcal{Y}$.[5][6] Since

$$\left\|\frac{w'}{e^T w'} - \frac{w''}{e^T w''}\right\| \leq \frac{\|w'-w''\|(e^T w'') + \|w''\|\,|e^T(w'-w'')|}{(e^T w')(e^T w'')} \leq 2(N_x+1)\|w'-w''\|$$

for any $w', w'' \in \mathcal{Q}^{N_x}$, we have that (24) is satisfied for all $w', w'' \in \mathcal{Q}^{N_x}$, $n \geq m \geq 0$ and any sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from Q, $\mathcal{Y}$. Hence, (i) is true.

Now, we shaw that (ii) is true, too. Let $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ be arbitrary sequences from Q, $\mathcal{Y}$ (respectively). As a consequence of Lemma 2, (i) and (9), we get

$$\|H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u,V)\| \leq C_{3,Q}\varepsilon_{1,Q}^{n-m}\|V\| + C_{2,Q}C_{3,Q}\sum_{i=m}^{n-1}\varepsilon_{1,Q}^{n-i-1} \leq C_{3,Q}\|V\| + C_{2,Q}C_{3,Q}(1-\varepsilon_{1,Q})^{-1}$$

[5] To deduce this, note that u, V, $y_{0:n}$, $G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{0:n}(u)$, $A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{0:n}(u)V$ have the same meaning respectively as quantities μ, $\tilde{\mu}$, y^n, $F_\theta^n(\mu, y^n)$, $\tilde{G}_\theta^n(\mu,\tilde{\mu},y^n)$ appearing in [36].

[6] Inequality (26) can also be obtained from [20, Theorem 2.1] or [23, Theorem 4.1]. Similarly, (24), (25) can be deduced from [19, Lemmas 3.4, 4.3, Proposition 5.2] (notice that $G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)$, $A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)$ have the same meaning respectively as $M_{m,n}$, $V[M_{m,n}, p_m]$ specified in [19, Section 5]).

for all $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq m \geq 0$. Due to the same arguments, we have

$$\begin{aligned}
&\|H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u',V') - H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u'',V'')\| \\
&\leq \|A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u') - A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u'')\|\|V'\| + \|A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u'')\|\|V'-V''\| \\
&\quad + \sum_{i=m}^{n-1} \|(\nabla_\theta G_{\vartheta_{i+1}})(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i}(u'),y_{i+1}) - (\nabla_\theta G_{\vartheta_{i+1}})(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i}(u''),y_{i+1})\|\|A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{i+1:n}(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i+1}(u'))\| \\
&\quad + \sum_{i=m}^{n-1} \|(\nabla_\theta G_{\vartheta_{i+1}})(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i}(u''),y_{i+1})\|\|A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{i+1:n}(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i+1}(u')) - A_{\boldsymbol{\vartheta},\boldsymbol{y}}^{i+1:n}(G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i+1}(u''))\| \\
&\leq C_{3,Q}\varepsilon_{1,Q}^{n-m}\|V'\|\|u'-u''\| + C_{3,Q}\varepsilon_{1,Q}^{n-m}\|V'-V''\| + C_{2,Q}C_{3,Q}\sum_{i=m}^{n-1}\varepsilon_{1,Q}^{n-i-1}\|G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i}(u') - G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i}(u'')\| \\
&\quad + C_{2,Q}C_{3,Q}\sum_{i=m}^{n-1}\varepsilon_{1,Q}^{n-i-1}\|G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i+1}(u') - G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:i+1}(u'')\| \\
&\leq C_{3,Q}\varepsilon_{1,Q}^{n-m}(\|u'-u''\|\|V'\| + \|V'-V''\|) + 2C_{2,Q}C_{3,Q}^2\varepsilon_{1,Q}^{n-m-1}(n-m)
\end{aligned}$$

for each $u', u'' \in \mathcal{P}^{N_x}$, $V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq m \geq 0$. Then, it is clear that there exist real numbers $\varepsilon_{2,Q} \in (0,1)$, $C_{4,Q} \in [1,\infty)$ such that (27), (28) hold for all $u, u', u'' \in \mathcal{P}^{N_x}$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$ and any sequence $\boldsymbol{\vartheta} = \{\vartheta\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from Q, $\mathcal{Y}$ (respectively). ■

Lemma 4: Suppose that Assumptions 3 and 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, there exists a real number $C_{5,Q} \in [1,\infty)$ such that

$$\|G_{\theta',\boldsymbol{y}}^{0:n}(u) - G_{\theta'',\boldsymbol{y}}^{0:n}(u)\| \leq C_{5,Q}\|\theta'-\theta''\|, \tag{29}$$

$$\|H_{\theta',\boldsymbol{y}}^{0:n}(u,V) - H_{\theta'',\boldsymbol{y}}^{0:n}(u,V)\| \leq C_{5,Q}\|\theta'-\theta''\|(1+\|V\|) \tag{30}$$

for all $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq 1$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from $\mathcal{Y}$.

Proof: Let $\tilde{C}_Q = C_{2,Q}C_{3,Q}C_{4,Q}^2$, while $\boldsymbol{y} = \{y_n\}_{n\geq 0}$ is an arbitrary sequence from $\mathcal{Y}$. It is straightforward to verify

$$G_{\theta',\boldsymbol{y}}^{0:n}(u) - G_{\theta'',\boldsymbol{y}}^{0:n}(u) = \sum_{i=0}^{n-1}\left(G_{\theta',\boldsymbol{y}}^{i:n}(G_{\theta'',\boldsymbol{y}}^{0:i}(u)) - G_{\theta',\boldsymbol{y}}^{i+1:n}(G_{\theta'',\boldsymbol{y}}^{0:i+1}(u))\right), \tag{31}$$

$$H_{\theta',\boldsymbol{y}}^{0:n}(u,V) - H_{\theta'',\boldsymbol{y}}^{0:n}(u,V) = \sum_{i=0}^{n-1}\left(H_{\theta',\boldsymbol{y}}^{i:n}(G_{\theta'',\boldsymbol{y}}^{0:i}(u), H_{\theta'',\boldsymbol{y}}^{0:i}(u,V)) - H_{\theta',\boldsymbol{y}}^{i+1:n}(G_{\theta'',\boldsymbol{y}}^{0:i+1}(u), H_{\theta'',\boldsymbol{y}}^{0:i+1}(u,V))\right) \tag{32}$$

for all $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq 0$. On the other side, Lemmas 2 and 3 yield

$$\begin{aligned}
\|G_{\theta',\boldsymbol{y}}^{i:n}(G_{\theta'',\boldsymbol{y}}^{0:i}(u)) - G_{\theta',\boldsymbol{y}}^{i+1:n}(G_{\theta'',\boldsymbol{y}}^{0:i+1}(u))\| &= \left\|G_{\theta',\boldsymbol{y}}^{i+1:n}\left(G_{\theta',\boldsymbol{y}}^{i:i+1}(G_{\theta'',\boldsymbol{y}}^{0:i}(u))\right) - G_{\theta',\boldsymbol{y}}^{i+1:n}\left(G_{\theta'',\boldsymbol{y}}^{i:i+1}(G_{\theta'',\boldsymbol{y}}^{0:i}(u))\right)\right\| \\
&\leq C_{3,Q}\varepsilon_{1,Q}^{n-i-1}\left\|G_{\theta',\boldsymbol{y}}^{i:i+1}(G_{\theta'',\boldsymbol{y}}^{0:i}(u)) - G_{\theta'',\boldsymbol{y}}^{i:i+1}(G_{\theta'',\boldsymbol{y}}^{0:i}(u))\right\| \\
&\leq \tilde{C}_Q\varepsilon_{1,Q}^{n-i-1}\|\theta'-\theta''\|
\end{aligned} \tag{33}$$

for any $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $0 \leq i < n$. Using the same lemmas, we also get

$$\begin{aligned}
&\|H^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)) - H^{i+1:n}_{\theta',\boldsymbol{y}}(G^{0:i+1}_{\theta'',\boldsymbol{y}}(u), H^{0:i+1}_{\theta'',\boldsymbol{y}}(u,V))\| \\
&= \Big\| H^{i+1:n}_{\theta',\boldsymbol{y}}\left(G^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)\right), H^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right) \\
&\quad - H^{i+1:n}_{\theta',\boldsymbol{y}}\left(G^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)\right), H^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right)\Big\| \\
&\leq C_{4,Q}\varepsilon^{n-i-1}_{2,Q}\left\|G^{i:i+1}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) - G^{i:i+1}_{\theta'',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u))\right\| \\
&\quad \cdot \left(1 + \left\|H^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right\| + \left\|H^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right\|\right) \\
&\quad + C_{4,Q}\varepsilon^{n-i-1}_{2,Q}\left\|H^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right) - H^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right\| \\
&\leq 3C_{2,Q}C^2_{4,Q}\varepsilon^{n-i-1}_{2,Q}\|\theta' - \theta''\|(1 + \|V\|) + C_{2,Q}C_{4,Q}\varepsilon^{n-i-1}_{2,Q}\|\theta' - \theta''\|(1 + \|H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\|) \\
&\leq 5\tilde{C}_Q\varepsilon^{n-i-1}_{2,Q}\|\theta' - \theta''\|(1 + \|V\|)
\end{aligned} \tag{34}$$

for each $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $0 \leq i < n$. Combining (31) – (34), we conclude that there exists a real number $C_{5,Q} \in [1, \infty)$ such that (29), (30) hold for all $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq 1$ and any sequence $\boldsymbol{y} = \{y_n\}_{n \geq 1}$ from $\mathcal{Y}$. ■

Lemma 5: Suppose that Assumptions 2 – 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, the following is true:

i) $f(\cdot)$ is well-defined and differentiable on Q.

ii) There exist real numbers $\varepsilon_{3,Q} \in (0,1)$, $C_{6,Q} \in [1, \infty)$ such that

$$\|(P^n F)(\theta, z) - \nabla f(\theta)\| \leq C_{6,Q}\varepsilon^n_{3,Q}(1 + \|V\|^2),$$

$$|(\Pi^n \phi)(\theta, \zeta) - f(\theta)| \leq C_{6,Q}\varepsilon^n_{3,Q}$$

for all $\theta \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$, $\zeta = (x, y, u) \in \mathcal{S}_\zeta$, $n \geq 1$.

Proof: Using [36, Theorems 4.1, 4.2] (with a few straightforward modifications), it can be deduced from Lemma 1 that there exist functions $g : \Theta \to \mathbb{R}^{d_\theta}$, $\psi : \Theta \to \mathbb{R}$ and real numbers $\varepsilon_{3,Q} \in (0,1)$, $C_{6,Q} \in [1, \infty)$ such that

$$\|(\tilde{P}^n \tilde{F})(\theta, z) - g(\theta)\| \leq C_{6,Q}\varepsilon^n_{3,Q}(1 + \|V\|^2), \tag{35}$$

$$|(\tilde{\Pi}^n \tilde{\phi})(\theta, \zeta) - \psi(\theta)| \leq C_{6,Q}\varepsilon^n_{3,Q} \tag{36}$$

for all $\theta \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$, $\zeta = (x, y, u) \in \mathcal{S}_\zeta$, $n \geq 1$.[7] Since $E|\phi_\theta(u_\theta, Y_1)| < \infty$ for any $\theta \in Q$ (due to Assumption 4), it follows from (12), (36) that $f(\cdot)$ is well-defined and identical to $\psi(\cdot)$ on Q. On the other side, Lemmas 1, 3 yield

$$\begin{aligned}
\|F_\theta(G^{0:n}_\theta(u, y_{1:n}), H^{0:n}_\theta(u, V, y_{1:n}), y_{n+1})\| &\leq C_{1,Q}\psi_Q(y_{n+1})(1 + \|H^{0:n}_\theta(u, V, y_{1:n})\|) \\
&\leq 2C_{1,Q}C_{4,Q}\psi_Q(y_{n+1})(1 + \|V\|)
\end{aligned}$$

[7]The same result can also be obtained from [20, Theorem 5.4]

for each $\theta \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from $\mathcal{Y}$. Then, Assumption 4 gives

$$E\left(\|F_\theta(G_\theta^{0:n}(u, Y_{1:n}), H_\theta^{0:n}(u, V, Y_{1:n}), Y_{n+1})\| \right| X_1 = x, Y_1 = y\big)$$
$$\leq 2C_{1,Q}C_{4,Q}(1 + \|V\|) \max_{x' \in \mathcal{X}} \int \psi_Q(y')Q(dy'|x') < \infty$$

for all $\theta \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Consequently, the dominated convergence theorem and (8), (10), (11) imply

$$\begin{aligned}
\nabla_\theta(\Pi^{n-1}\phi)(\theta, \zeta) =& E\left(\nabla_\theta \left(\phi_\theta(G_\theta^{0:n}(u, Y_{1:n}), Y_{n+1}) \right) \right| X_1 = x, Y_1 = y\big) \\
=& E\left(F_\theta(G_\theta^{0:n}(u, Y_{1:n}), H_\theta^{0:n}(u, 0, Y_{1:n}), Y_{n+1}) \right| X_1 = x, Y_1 = y\big) \\
=& (P^{n-1}F)(\theta, (\zeta, 0))
\end{aligned} \tag{37}$$

for any $\theta \in Q$, $\zeta = (x, y, u) \in \mathcal{S}_\zeta$, $n > 1$ (here, 0 stands for $d_\theta \times N_x$ zero matrix). As $(\Pi^n\phi)(\theta, \zeta)$ and $(P^nF)(\theta, z)$ converge (respectively) to $\psi(\theta)$ and $g(\theta)$ uniformly in $\theta \in Q$ for each $z \in \mathcal{S}_z$, $\zeta \in \mathcal{S}_\zeta$ (due to (10), (11), (35), (36)), it follows from (37) that Part (i) is true. Part (ii) is then a direct consequence of (10), (11), (35), (36). ∎

Lemma 6: Suppose that Assumptions 2 – 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, there exists a real number $C_{7,Q} \in [1, \infty)$ such that

$$\|(P^nF)(\theta', z) - (P^nF)(\theta'', z)\| \leq C_{7,Q}\|\theta' - \theta''\|(1 + \|V\|^2) \tag{38}$$

for all $\theta', \theta'' \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$, $n \geq 1$.

Proof: Let Owing to Lemmas 1, 3 and 4, we have

$$\begin{aligned}
&\|F_{\theta'}(G_{\theta'}^{0:n}(u, y_{1:n}), H_{\theta'}^{0:n}(u, V, y_{1:n}), y_{n+1}) - F_{\theta''}(G_{\theta''}^{0:n}(u, y_{1:n}), H_{\theta''}^{0:n}(u, V, y_{1:n}), y_{n+1})\| \\
\leq& C_{1,Q}\psi_Q(y_{n+1})(1 + \|H_{\theta'}^{0:n}(u, V, y_{1:n})\| + \|H_{\theta''}^{0:n}(u, V, y_{1:n})\|) \\
&\cdot (\|\theta' - \theta''\| + \|G_{\theta'}^{0:n}(u, y_{1:n}) - G_{\theta''}^{0:n}(u, y_{1:n})\| + \|H_{\theta'}^{0:n}(u, V, y_{1:n}) - H_{\theta''}^{0:n}(u, V, y_{1:n})\|) \\
\leq& 9C_{1,Q}C_{4,Q}C_{5,Q}\psi_Q(y_{n+1})(1 + \|V\|)^2\|\theta' - \theta''\|
\end{aligned}$$

for all $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq 1$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from $\mathcal{Y}$. Consequently,

$$\begin{aligned}
\|(\Pi^nF)(\theta', z) - (\Pi^nF)(\theta'', z)\| \leq& E\big(\|F_{\theta'}(G_{\theta'}^{0:n}(u, y_{1:n}), H_{\theta'}^{0:n}(u, V, y_{1:n}), y_{n+1}) \\
&- F_{\theta''}(G_{\theta''}^{0:n}(u, y_{1:n}), H_{\theta''}^{0:n}(u, V, y_{1:n}), y_{n+1})\| \, |X_1 = x, Y_1 = y\big) \\
\leq& 9C_{1,Q}C_{4,Q}C_{5,Q}(1 + \|V\|)^2\|\theta' - \theta''\| \max_{x' \in \mathcal{X}} \int \psi_Q(y')Q(dy'|x')
\end{aligned}$$

for each $\theta', \theta'' \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$. Then, it can be deduced from Assumption 4 that there exists a real number $C_{7,Q} \in [1, \infty)$ such that (38) holds for all $\theta', \theta'' \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$. ∎

Lemma 7: Suppose that Assumptions 3 and 4 hold. Let $Q \subset \Theta$ be an arbitrary compact set. Then, there exist real numbers $\delta_{3,Q}, \varepsilon_{4,Q} \in (0, 1)$, $C_{8,Q} \in [1, \infty)$ such that the following is true:

i) $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ is analytical in (η, w) on $V_{\delta_{3,Q}}(Q) \times V_{\delta_{3,Q}}(\mathcal{P}^{N_x})$ for each $n \geq 0$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ from $\mathcal{Y}$.

ii) Inequalities

$$d(\hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w),\mathcal{P}^{N_x}) \leq \min\{\delta_Q,\delta_{1,Q},\delta_{2,Q}\},$$

$$\|\hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w')-\hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w'')\| \leq C_{8,Q}\varepsilon^n_{4,Q}\|w'-w''\|$$

hold for all $\eta \in V_{\delta_{3,Q}}(Q)$, $w,w',w'' \in V_{\delta_{3,Q}}(\mathcal{P}^{N_x})$ and any sequence $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ from $\mathcal{Y}$ (δ_Q is specified in Assumption 4).

Proof: Let $\boldsymbol{y}=\{y_n\}_{n\geq 1}$ be an arbitrary sequence from $\mathcal{Y}$. Moreover, let $k_Q=\min\{n\geq 1 : C_{3,Q}\varepsilon^n_{1,Q} \leq \varepsilon_{1,Q}/2\}$, while $\tilde{\delta}_{1,Q}=\min\{\delta_Q,\delta_{1,Q},\delta_{2,Q}\}$, $\tilde{\delta}_{2,Q}=4^{-k_Q}C^{-k_Q}_{2,Q}\tilde{\delta}_{1,Q}$.

First, we prove by induction (in k) that

$$d(\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w),\mathcal{P}^{N_x}) \leq (2^{k+1}C^k_{2,Q}-1)\tilde{\delta}_{2,Q} \leq \tilde{\delta}_{1,Q} \tag{39}$$

for all $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$, $0 \leq k \leq k_Q$. Obviously, (39) is true when $k=0$, $n\geq 0$, $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$. Suppose now that (39) holds for each $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$ and some $0 \leq k < k_Q$. Then, Lemma 2 implies

$$\begin{aligned}\|\hat{G}^{n:n+k+1}_{\eta,\boldsymbol{y}}(w)-G_\theta(u,y_{n+k+1})\| =&\|\hat{G}_\eta(\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w),y_{n+k+1})-\hat{G}_\theta(u,y_{n+k+1})\|\\ \leq& C_{2,Q}(\|\eta-\theta\|+\|\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w)-u\|)\end{aligned}$$

for any $\theta \in Q$, $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $u \in \mathcal{P}^{N_x}$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$. Therefore,

$$\begin{aligned}d(\hat{G}^{n:n+k+1}_{\eta,\boldsymbol{y}}(w),\mathcal{P}^{N_x}) \leq& C_{2,Q}\left(d(\eta,Q)+d(\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w),\mathcal{P}^{N_x})\right)\\ \leq& 2^{k+1}C^{k+1}_{2,Q}\tilde{\delta}_{2,Q}\\ \leq& (2^{k+2}C^{k+1}_{2,Q}-1)\tilde{\delta}_{2,Q} \leq \tilde{\delta}_{1,Q}\end{aligned}$$

for any $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$. Hence, (39) is satisfied for all $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$, $0 \leq k \leq k_Q$.

Let $\tilde{\delta}_{3,Q}=\tilde{\delta}_{2,Q}/2$. Since $\hat{G}^{n:n}_{\eta,\boldsymbol{y}}(w)=w$ and $\hat{G}^{n:n+k+1}_{\eta,\boldsymbol{y}}(w)=\hat{G}_\eta(\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w),y_{n+k+1})$, it can be deduced from Assumption 4 and (39) that $\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w)$ is analytic in (η,w) on $V_{\tilde{\delta}_{2,Q}}(Q)\times V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$ for each $n \geq 0$, $0 \leq k \leq k_Q$ (notice that a composition of two analytic functions is analytic, too). Due to Assumption 4 and (39), we also have

$$\|\hat{G}^{n:n+k+1}_{\eta,\boldsymbol{y}}(w)\| = \|\hat{G}_\eta(\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w),y_{n+k+1})\| \leq K_Q \tag{40}$$

for all $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$, $0 \leq k \leq k_Q$ (K_Q is defined in Assumption 4). As a consequence of Cauchy inequality for analytic functions and (40), there exists a real number $\tilde{C}_{1,Q} \in [1,\infty)$ depending exclusively on K_Q, d_θ, N_x ($\tilde{C}_{1,Q}$ can be selected as $\tilde{C}_{1,Q}=4(d_\theta+N_x)K_Q/\tilde{\delta}^2_{2,Q}$) such that

$$\max\{\|\nabla_{(\eta,w)}\hat{G}^{n:n+k}_{l,\eta,\boldsymbol{y}}(w)\|,\|\nabla^2_{(\eta,w)}\hat{G}^{n:n+k}_{l,\eta,\boldsymbol{y}}(w)\|\} \leq \tilde{C}_{1,Q}$$

for any $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$, $0 \leq k \leq k_Q$, $1 \leq l \leq N_x$ ($\hat{G}^{n:n+k}_{l,\eta,\boldsymbol{y}}(w)$ denote the l-th component of $\hat{G}^{n:n+k}_{\eta,\boldsymbol{y}}(w)$). Consequently, there exists another real number $\tilde{C}_{2,Q} \in [1\infty)$ depending exclusively on K_Q, d_θ,

N_x such that

$$\max\{\|\hat{G}^{n:n+k}_{\eta',\boldsymbol{y}}(w') - \hat{G}^{n:n+k}_{\eta'',\boldsymbol{y}}(w'')\|, \|\nabla_w \hat{G}^{n:n+k}_{\eta',\boldsymbol{y}}(w') - \nabla_w \hat{G}^{n:n+k}_{\eta'',\boldsymbol{y}}(w'')\|\} \\ \leq \tilde{C}_{2,Q}(\|\eta' - \eta''\| + \|w' - w''\|) \tag{41}$$

for each $\eta', \eta'' \in V_{\tilde{\delta}_{3,Q}}(Q)$, $w', w'' \in V_{\tilde{\delta}_{3,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$, $0 \leq k \leq k_Q$.

Let $\tilde{\delta}_{4,Q} = \min\{\tilde{\delta}_{3,Q}, 4^{-1}\tilde{C}^{-1}_{2,Q}\varepsilon_{1,Q}\}$. Owing to Lemma 3 (Part (i)), we have

$$\|G^{n:n+k_Q}_{\theta,\boldsymbol{y}}(u') - G^{n:n+k_Q}_{\theta,\boldsymbol{y}}(u'')\| \leq C_{3,Q}\varepsilon^{k_Q}_{1,Q}\|u' - u''\| \leq (\varepsilon_{1,Q}/2)\|u' - u''\|$$

for all $\theta \in Q$, $u', u'' \in [0,\infty)^{N_x} \setminus \{0\}$, $n \geq 0$. Therefore, $\|\nabla_u G^{n:n+k_Q}_{\theta,\boldsymbol{y}}(w)\| \leq \varepsilon_{1,Q}/2$ for each $\theta \in Q$, $u \in [0,\infty)^{N_x} \setminus \{0\}$, $n \geq 0$, which, together with (41) yields

$$\begin{aligned}\|\nabla_w \hat{G}^{n:n+k_Q}_{\eta,\boldsymbol{y}}(w)\| \leq& \|\nabla_u G^{n:n+k_Q}_{\theta,\boldsymbol{y}}(u)\| + \|\nabla_w \hat{G}^{n:n+k_Q}_{\eta,\boldsymbol{y}}(w) - \nabla_w \hat{G}^{n:n+k_Q}_{\theta,\boldsymbol{y}}(u)\| \\ \leq& \varepsilon_{1,Q}/2 + \tilde{C}_{2,Q}(\|\theta - \eta\| + \|u - w\|)\end{aligned}$$

for any $\theta \in Q$, $\eta \in V_{\tilde{\delta}_{3,Q}}(Q)$, $u \in \mathcal{P}^{N_x}$, $w \in V_{\tilde{\delta}_{3,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$. Consequently,

$$\|\nabla_w \hat{G}^{n:n+k_Q}_{\eta,\boldsymbol{y}}(w)\| \leq \varepsilon_{1,Q}/2 + \tilde{C}_{2,Q}(d(\eta, Q) + d(w, \mathcal{P}^{N_x})) \leq \varepsilon_{1,Q}$$

for each $\eta \in V_{\tilde{\delta}_{4,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$. Thus,

$$\|\hat{G}^{n:n+k_Q}_{\eta,\boldsymbol{y}}(w') - \hat{G}^{n:n+k_Q}_{\eta,\boldsymbol{y}}(w'')\| \leq \int_0^1 \|\nabla_w \hat{G}^{n:n+k_Q}_{\eta,\boldsymbol{y}}(tw' + (1-t)w'')\|\|w' - w''\|dt \leq \varepsilon_{1,Q}\|w' - w''\| \tag{42}$$

for all $\eta \in V_{\tilde{\delta}_{4,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$.

Let $\tilde{\delta}_{5,Q} = (1 - \varepsilon_{1,Q})\tilde{\delta}_{4,Q}\tilde{C}^{-1}_{2,Q}$. Now, we prove by induction (in i) that

$$d(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w), \mathcal{P}^{N_x}) \leq \tilde{\delta}_{4,Q} \tag{43}$$

for each $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$, $i \geq 0$. Obviously, (43) is true when $i = 0$, $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$. Suppose that (43) holds for all $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$ and some $i \geq 0$. Then, (41), (42) imply

$$\begin{aligned}\|\hat{G}^{0:(i+1)k_Q}_{\eta,\boldsymbol{y}}(w) - G^{ik_Q:(i+1)k_Q}_{\theta,\boldsymbol{y}}(u)\| \leq& \|\hat{G}^{ik_Q:(i+1)k_Q}_{\eta,\boldsymbol{y}}(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w)) - \hat{G}^{ik_Q:(i+1)k_Q}_{\eta,\boldsymbol{y}}(u)\| \\ &+ \|\hat{G}^{ik_Q:(i+1)k_Q}_{\eta,\boldsymbol{y}}(u) - \hat{G}^{ik_Q:(i+1)k_Q}_{\theta,\boldsymbol{y}}(u)\| \\ \leq& \varepsilon_{1,Q}\|\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w) - u\| + \tilde{C}_{2,Q}\|\theta - \eta\|\end{aligned}$$

for any $\theta \in Q$, $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $u \in \mathcal{P}^{N_x}$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$. Therefore,

$$d(\hat{G}^{0:(i+1)k_Q}_{\eta,\boldsymbol{y}}(w), \mathcal{P}^{N_x}) \leq \varepsilon_{1,Q} d(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w), \mathcal{P}^{N_x}) + \tilde{C}_{2,Q} d(\eta, Q) \leq \varepsilon_{1,Q}\tilde{\delta}_{4,Q} + \tilde{C}_{2,Q}\tilde{\delta}_{5,Q} = \tilde{\delta}_{4,Q}$$

for each $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$. Hence, (43) holds for all $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$, $i \geq 0$.

Let $\delta_{3,Q} = \min\{\tilde{\delta}_{4,Q}, \tilde{\delta}_{5,Q}\}$. As $\hat{G}^{0:0}_{\eta,\boldsymbol{y}}(w) = w$ and $\hat{G}^{0:(i+1)k_Q}_{\eta,\boldsymbol{y}}(w) = \hat{G}^{ik_Q:(i+1)k_Q}_{\eta,\boldsymbol{y}}(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w))$, it can be deduced from (43) that $\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w)$ is analytical in (η, w) on $V_{\tilde{\delta}_{5,Q}}(Q) \times V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$ for each $i \geq 0$ (notice that $\hat{G}^{ik_Q:(i+1)k_Q}_{\eta,\boldsymbol{y}}(w)$ is analytic in (η, w) on $V_{\tilde{\delta}_{5,Q}}(Q) \times V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$ for any $i \geq 0$). Since $\hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w) = \hat{G}^{ik_Q:n}_{\eta,\boldsymbol{y}}(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w))$ for $i = \lfloor n/k_Q \rfloor$, we conclude from (43) that $\hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w)$ is analytical in (η, w) on $V_{\tilde{\delta}_{5,Q}}(Q) \times V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x}) \supseteq$

$V_{\delta_{3,Q}}(Q)\times V_{\delta_{3,Q}}(\mathcal{P}^{N_x})$ for all $n\geq 0$ (notice that $\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:ik_Q+j}(w)$ is analytical in (η,w) on $V_{\tilde{\delta}_{5,Q}}(Q)\times V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$ for any $i\geq 0$, $0\leq j\leq k_Q$). On the other side, (39), (43) yield

$$d(\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w),\mathcal{P}^{N_x})=d(\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:n}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w)),\mathcal{P}^{N_x})\leq\tilde{\delta}_{1,Q}=\min\{\delta_Q,\delta_{1,Q},\delta_{2,Q}\} \tag{44}$$

for all $\eta\in V_{\tilde{\delta}_{5,Q}}(Q)\supseteq V_{\delta_{3,Q}}(Q)$, $w\in V_{\tilde{\delta}_{5,Q}}(\mathcal{P}^{N_x})\supseteq V_{\delta_{3,Q}}(\mathcal{P}^{N_x})$, $n\geq 0$ and $i=\lfloor n/k_Q\rfloor$.

Let $\varepsilon_{4,Q}=\varepsilon_{1,Q}^{1/k_Q}$, $C_{8,Q}=\tilde{C}_{2,Q}\varepsilon_{1,Q}^{-1}$. Owing to (42), (43), we have

$$\begin{aligned}\|\hat{G}_{\eta,\boldsymbol{y}}^{0:(i+1)k_Q}(w')-\hat{G}_{\eta,\boldsymbol{y}}^{0:(i+1)k_Q}(w'')\|=&\|\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:(i+1)k_Q}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w'))-\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:(i+1)k_Q}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w''))\|\\ \leq&\varepsilon_{1,Q}\|\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w')-\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w'')\|\end{aligned}$$

for any $\eta\in V_{\tilde{\delta}_{5,Q}}(Q)$, $w',w''\in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$, $i\geq 0$. Therefore,

$$\|\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w')-\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w'')\|\leq\varepsilon_{1,Q}^{i}\|w'-w''\|$$

for each $\eta\in V_{\tilde{\delta}_{5,Q}}(Q)$, $w',w''\in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$, $i\geq 0$. Consequently, (41), (43) yield

$$\begin{aligned}\|\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w')-\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w'')\|=&\|\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:n}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w'))-\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:n}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w''))\|\\ \leq&\tilde{C}_{2,Q}\|\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w')-\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w'')\|\\ \leq&\tilde{C}_{2,Q}\varepsilon_{1,Q}^{i}\|w'-w''\|\\ \leq&C_{8,Q}\varepsilon_{4,Q}^{n}\|w'-w''\|\end{aligned}$$

for each $\eta\in V_{\tilde{\delta}_{5,Q}}(Q)\supseteq V_{\delta_{3,Q}}(Q)$, $w',w''\in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})\supseteq V_{\delta_{3,Q}}(\mathcal{P}^{N_x})$, $n\geq 0$, $i=\lfloor n/k_Q\rfloor$ (notice that $\tilde{C}_{2,Q}\varepsilon_{1,Q}^{i}=\tilde{C}_{2,Q}\varepsilon_{4,Q}^{-(n-ik_Q)}\varepsilon_{4,Q}^{n}\leq C_{8,Q}\varepsilon_{4,Q}^{n}$). Then, it is clear that $\delta_{3,Q}$, $\varepsilon_{4,Q}$, $C_{8,Q}$ meet the requirements of the lemma. ■

B. Analyticity

In this subsection, using the results of the Subsection IV-A (Lemma 7), the analyticity of the objective function $f(\cdot)$ is shown and Theorem 1 is proved. The proof is based on the analytic continuation techniques and the methods developed in [13].

Proof of Theorem 1: Let

$$\hat{\psi}_{\eta}^{n}(w,x)=E\left(\hat{\phi}_{\eta}(\hat{G}_{\eta}^{0:n}(w,Y_{1:n}),Y_{n+1})\middle|X_1=x\right)$$

for $\eta\in\mathbb{C}^{d_\theta}$, $w\in\mathbb{C}^{N_x}$, $x\in\mathcal{X}$, $n\geq 1$. Then, using (7), it is straightforward to verify

$$\begin{aligned}\hat{\psi}_{\eta}^{n+1}(w,x)=&E\left(E\left(\hat{\phi}_{\eta}(\hat{G}_{\eta}^{0:n}(\hat{G}_{\eta}(w,Y_1),Y_{2:n+1}),Y_{n+2})\middle|X_1,X_2,Y_1\right)\middle|X_1=x\right)\\ =&E(\hat{\psi}_{\eta}^{n}(\hat{G}_{\eta}(w,Y_1),X_2)|X_1=x)\end{aligned} \tag{45}$$

for each $\eta \in \mathbb{C}^{d_\theta}$, $w \in \mathbb{C}^{N_x}$, $x \in \mathcal{X}$, $n \geq 0$. It is also easy to show

$$\begin{aligned}
&\hat{\psi}_\eta^n(w', x') - \hat{\psi}_\eta^n(w'', x'') \\
=&E\left(\hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w', Y_{1:n}), Y_{n+1}) - \hat{\phi}_\eta(\hat{G}_\eta^{0:n}(e_0, Y_{1:n}), Y_{n+1}) \Big| X_1 = x' \right) \\
&- E\left(\hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w'', Y_{1:n}), Y_{n+1}) - \hat{\phi}_\eta(\hat{G}_\eta^{0:n}(e_0, Y_{1:n}), Y_{n+1}) \Big| X_1 = x'' \right) \\
&+ \sum_{k=1}^{n-1} \sum_{x \in \mathcal{X}} E\left(\hat{\phi}_\eta(\hat{G}_\eta^{0:n-k+1}(e_0, Y_{k:n}), Y_{n+1}) - \hat{\phi}_\eta(\hat{G}_\eta^{0:n-k}(e_0, Y_{k+1:n}), Y_{n+1}) \Big| X_k = x \right) \\
&\qquad\qquad \cdot (p^{k-1}(x|x') - \pi(x)) \\
&- \sum_{k=1}^{n-1} \sum_{x \in \mathcal{X}} E\left(\hat{\phi}_\eta(\hat{G}_\eta^{0:n-k+1}(e_0, Y_{k:n}), Y_{n+1}) - \hat{\phi}_\eta(\hat{G}_\eta^{0:n-k}(e_0, Y_{k+1:n}), Y_{n+1}) \Big| X_k = x \right) \\
&\qquad\qquad \cdot (p^{k-1}(x|x'') - \pi(x)) \\
&+ \sum_{x \in \mathcal{X}} E(\hat{\phi}_\eta(\hat{G}_\eta(e_0, Y_n), Y_{n+1}) | X_n = x)(p^{n-1}(x|x') - \pi(x)) \\
&- \sum_{x \in \mathcal{X}} E(\hat{\phi}_\eta(\hat{G}_\eta(e_0, Y_n), Y_{n+1}) | X_n = x)(p^{n-1}(x|x') - \pi(x))
\end{aligned} \tag{46}$$

for all $\eta \in \mathbb{C}^{d_\theta}$, $w', w'' \in \mathbb{C}^{N_x}$, $x', x'' \in \mathcal{X}$, $n \geq 1$, where $e_0 = [1 \cdots 1]^T / N_x \in \mathbb{R}^{N_x}$ and $p^{k-1}(x'|x) = P(X_k = x'|X_1 = x)$, $\pi(x) = \lim_{k \to \infty} P(X_k = x)$. On the other side, Assumption 2 implies that $\pi(\cdot)$ is well-defined and that there exist real numbers $\tilde{\varepsilon} \in (0, 1)$, $\tilde{C} \in [1, \infty)$ such that

$$|p^n(x'|x) - \pi(x')| \leq \tilde{C} \tilde{\varepsilon}^n \tag{47}$$

for each $x, x' \in \mathcal{X}$, $n \geq 0$.

Let $Q \subset \Theta$ be an arbitrary compact set, while $\tilde{\delta}_{1,Q} = \min\{\delta_Q, \delta_{1,Q}, \delta_{2,Q}, \delta_{3,Q}\}$, $\tilde{\delta}_{2,Q} = \tilde{\delta}_{1,Q}/2$. Owing to Assumption 4 and Lemma 7, $\hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w, y_{1:n}), y_{n+1})$ is analytic in (η, w) on $V_{\tilde{\delta}_{1,Q}}(Q) \times V_{\tilde{\delta}_{1,Q}}(\mathcal{P}^{N_x})$ for each $n \geq 0$ and any sequence $\boldsymbol{y} = \{y_n\}_{n \geq 1}$ from $\mathcal{Y}$. Due to Assumption 4 and Lemma 7, we also have

$$|\hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w, y_{1:n}), y_{n+1})| \leq \psi_Q(y_{n+1})$$

for all $\eta \in V_{\tilde{\delta}_{1,Q}}(Q)$, $w \in V_{\tilde{\delta}_{1,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$ and any sequence $\boldsymbol{y} = \{y_n\}_{n \geq 1}$ from $\mathcal{Y}$. Consequently, Cauchy inequality for analytic functions implies that there exists a real number $\tilde{C}_{1,Q} \in [1, \infty)$ such that

$$\|\nabla_\eta \hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w, y_{1:n}), y_{n+1})\| \leq \tilde{C}_{1,Q} \psi_Q(y_{n+1}) \tag{48}$$

for each $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$ and any sequence $\boldsymbol{y} = \{y_n\}_{n \geq 1}$ from $\mathcal{Y}$. Since

$$E(\psi_Q(Y_{n+1}) | X_1 = x) \leq \max_{x' \in \mathcal{X}} \int \psi_Q(y') Q(dy'|x') < \infty \tag{49}$$

for all $x \in \mathcal{X}$, $n \geq 0$, it follows from the dominated convergence theorem and (48) that $\hat{\psi}_\eta^n(w, x)$ is differentiable (and thus, analytic) in η on $V_{\tilde{\delta}_{2,Q}}(Q)$ for any $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 0$.

Let $\tilde{\varepsilon}_Q = \max\{\varepsilon_{4,Q}, \tilde{\varepsilon}\}$. Due to Lemmas 1 and 7, we have

$$|\hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w', y_{1:n}), y_{n+1}) - \hat{\phi}_\eta(\hat{G}_\eta^{0:n}(w'', y_{1:n}), y_{n+1})| \leq C_{1,Q} C_{8,Q} \varepsilon_{4,Q}^n \psi_Q(y_{n+1}) \|w' - w''\|, \tag{50}$$

$$\begin{aligned} &|\hat{\phi}_\eta(\hat{G}_\eta^{0:n-k+1}(w, y_{k:n}), y_{n+1}) - \hat{\phi}_\eta(\hat{G}_\eta^{0:n-k}(w, y_{k+1:n}), y_{n+1})| \\ &\quad \leq C_{1,Q} \psi_Q(y_{n+1}) \|\hat{G}_\eta^{0:n-k}(\hat{G}_\eta(w, y_k), y_{k+1:n}) - \hat{G}_\eta^{0:n-k}(w, y_{k+1:n})\| \\ &\quad \leq C_{1,Q} C_{8,Q} \varepsilon_{4,Q}^{n-k} \psi_Q(y_{n+1}) \|\hat{G}_\eta(w, y_k) - w\| \end{aligned} \tag{51}$$

for each $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w, w', w'' \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 1$, $0 < k \leq n$ and any sequence $\boldsymbol{y} = \{y_n\}_{n \geq 1}$ from $\mathcal{Y}$. Using (47), (49) – (51), we deduce that there exists a real number $\tilde{C}_{2,Q} \in [1, \infty)$ such that the absolute value of the each term on right-hand side of (46) is bounded by $\tilde{C}_{2,Q} \tilde{\varepsilon}_Q^n$ for any $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w', w'' \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $x, x' \in \mathcal{X}$, $n \geq 1$. Therefore,

$$|\hat{\psi}_\eta^n(w', x') - \hat{\psi}_\eta^n(w'', x'')| \leq 2\tilde{C}_{2,Q} \tilde{\varepsilon}_Q^n (n+1) \tag{52}$$

for all $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w', w'' \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $x', x'' \in \mathcal{X}$, $n \geq 1$. Consequently, (45) yields

$$|\hat{\psi}_\eta^{n+1}(w, x) - \hat{\psi}_\eta^n(w, x)| \leq E\left(|\hat{\psi}_\eta^n(\hat{G}_\eta(w, Y_1), X_2) - \hat{\psi}_\eta^n(w, x)| \,\Big|\, X_1 = x \right) \leq 2\tilde{C}_{2,Q} \tilde{\varepsilon}_Q^n (n+1) \tag{53}$$

for each $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $x \in \mathcal{X}$, $n \geq 1$. Owing to (52), (53), there exists a function $\hat{\psi} : \mathbb{C}^{d_\theta} \to \mathbb{C}$ such that $\hat{\psi}_\eta^n(w, x)$ converges to $\hat{\psi}(\eta)$ uniformly in $(\eta, w, x) \in V_{\tilde{\delta}_{2,Q}}(Q) \times V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x}) \times \mathcal{X}$. As the uniform limit of analytic functions is also an analytic function (see [37, Theorem 2.4.1]), $\hat{\psi}(\cdot)$ is analytic on $V_{\tilde{\delta}_{2,Q}}(Q)$. On the other side, since

$$\hat{\phi}_\theta^n(u, x) = E\left(\phi_\theta(G_\theta^{0:n}(u, Y_{1:n}), Y_{n+1}) \big|\, X_1 = x \right) = E\left((\Pi^{n-1}\phi)(\theta, (x, Y_1, u)) \big|\, X_1 = x \right)$$

for all $\theta \in \Theta$, $u \in \mathcal{P}^{N_x}$, $x \in \mathcal{X}$, $n \geq 1$, Lemma 5 implies $f(\theta) = \hat{\psi}(\theta)$ for any $\theta \in Q$. Then, it is clear that Part (i) is true, while Part (ii) follows from the Lojasiewicz inequality (see e.g., [17], [26], [27]) and the analyticity of $f(\cdot)$. ■

As a direct consequence of [17, Theorem ŁI, Page 775] and Theorem 1, we have the following corollary:

Corollary 5: Let Assumptions 2 – 4 hold. Then, for any compact set $Q \subset \Theta$ and real number $a \in f(Q)$, there exist real numbers $\delta_{Q,a} \in (0, 1)$, $\mu_{Q,a} \in (1, 2]$, $M_{Q,a} \in [1, \infty)$ such that

$$|f(\theta) - a| \leq M_{Q,a} \|\nabla f(\theta)\|^{\mu_{Q,a}}$$

for all $\theta \in Q$ satisfying $|f(\theta) - a| \leq \delta_{Q,a}$.

Remark: Obviously, if $Q \subseteq \{\theta' \in \mathbb{R}^{d_\theta} : \|\theta' - \theta\| \leq \delta_\theta\}$ and $a = f(\theta)$ for some $\theta \in \mathbb{R}^{d_\theta}$, then $\mu_{Q,a}$ and $M_{Q,a}$ can be selected as $\mu_{Q,a} = \mu_\theta$ and $M_{Q,a} = M_\theta$ (δ_θ, μ_θ, M_θ are specified in the statement of Theorem 1).

C. *Decomposition of Algorithm (1) – (3)*

Relying on the results of Subsection IV-A (Lemmas 1 – 6), equivalent representations of recursion (1) – (3) and their asymptotic properties are analyzed in this subsection. The analysis is based on the techniques developed in [2, Part II]. The results of this subsection are a crucial prerequisite for the analysis carried out in the next subsection.

In this subsection, the following notation is used. For $n \geq 0$, let $Z_{n+1} = (X_{n+1}, Y_{n+1}, U_n, V_n)$, while

$$\xi_n = F(\theta_n, Z_{n+1}) - \nabla f(\theta_n),$$
$$\phi'_n = \alpha_n (\nabla f(\theta_n))^T \xi_n,$$
$$\phi''_n = \int_0^1 (\nabla f(\theta_n + t(\theta_{n+1} - \theta_n)) - \nabla f(\theta_n))^T (\theta_{n+1} - \theta_n) dt$$

and $\phi_n = \phi'_n + \phi''_n$ ($F(\cdot, \cdot)$ is defined in the beginning of the previous subsection). Then, algorithm (1) – (3) admits the following representations:

$$\begin{aligned}\theta_{n+1} =& \theta_n + \alpha_n F(\theta_n, Z_{n+1}) \\ =& \theta_n + \alpha_n (\nabla f(\theta_n) + \xi_n), \quad n \geq 0.\end{aligned}$$

Moreover, we have

$$f(\theta_{n+1}) = f(\theta_n) + \alpha_n \|\nabla f(\theta_n)\|^2 + \phi_n$$

for $n \geq 0$. We also conclude

$$P(Z_{n+1} \in B | \theta_0, Z_0, \ldots, \theta_n, Z_n) = P_{\theta_n}(Z_n, B)$$

w.p.1 for $n \geq 0$ and any Borel-measurable set $B \subseteq \mathcal{S}_z$ ($P_\theta(\cdot, \cdot)$ is also introduced in the beginning of the previous subsection).

Lemma 8: Suppose that Assumptions 2 – 4 hold. Then, there exists a Borel-measurable function $\Phi : \Theta \times \mathcal{S}_z \to \mathbb{R}^{d_\theta}$ with the following properties:

i) $\Phi(\theta, \cdot)$ is integrable with respect to $P_\theta(z, \cdot)$ and

$$F(\theta, z) - \nabla f(\theta) = \Phi(\theta, z) - (P\Phi)(\theta, z) \tag{54}$$

for all $\theta \in \Theta$, $z \in \mathcal{S}_z$.

ii) For any compact set $Q \subset \Theta$ and a real number $s \in (0, 1)$, there exists a Borel-measurable function $\varphi_{Q,s} : \mathcal{S}_z \to [1, \infty)$ such that

$$\max\{\|F(\theta, z)\|, \|\Phi(\theta, z)\|, \|(P\Phi)(\theta, z)\|\} \leq \varphi_{Q,s}(z), \tag{55}$$

$$\|(P\Phi)(\theta', z) - (P\Phi)(\theta'', z)\| \leq \varphi_{Q,s}(z) \|\theta' - \theta''\|^s, \tag{56}$$

$$\sup_{n \geq 0} E\left(\varphi^2_{Q,s}(Z_n) I_{\{\tau_Q \geq n\}} \middle| Z_0 = z \right) < \infty \tag{57}$$

for all $\theta, \theta', \theta'' \in Q$, $z \in \mathcal{S}_z$, where

$$\tau_Q = \inf\{n \geq 0 : \theta_n \notin Q\}.$$

Proof: Let $Q \subseteq \Theta$ be an arbitrary compact set. Owing to Lemmas 1 and 5, there exists a real number $\tilde{C}_{1,Q} \in [1, \infty)$ such that

$$\sum_{k=0}^{\infty} \|(P^k F)(\theta, z) - \nabla f(\theta)\| \leq \tilde{C}_{1,Q} \psi_Q(y)(1 + \|V\|^2) \tag{58}$$

for all $\theta \in Q$, $z = (x, y, u, v) \in \mathcal{S}_z$ ($(P^0 F)(\theta, z)$ stands for $F(\theta, z)$). Consequently, $\sum_{k=0}^{\infty}((P^k F)(\theta, z) - \nabla f(\theta))$ if well-defined and finite for each $\theta \in Q$, $z \in \mathcal{S}_z$. We also have

$$\begin{aligned}
&\left\| \sum_{k=1}^{\infty}((P^k F)(\theta', z) - \nabla f(\theta')) - \sum_{k=1}^{\infty}((P^k F)(\theta'', z) - \nabla f(\theta'')) \right\| \\
\leq & \sum_{k=1}^{n} \left\| (P^k F)(\theta', z) - (P^k F)(\theta'', z) \right\| + n \left\| \nabla f(\theta') - \nabla f(\theta'') \right\| \\
&+ \sum_{k=n+1}^{\infty} \left\| (P^k F)(\theta', z) - \nabla f(\theta') \right\| + \sum_{k=n+1}^{\infty} \left\| (P^k F)(\theta'', z) - \nabla f(\theta'') \right\|
\end{aligned}$$

for each $\theta', \theta'' \in \Theta$, $z \in \mathcal{S}_z$, $n \geq 1$. Then, using Lemmas 5 and 6, it can be deduced that there exist real numbers $\tilde{\varepsilon}_Q \in (0, 1)$, $\tilde{C}_{2,Q} \in [1, \infty)$ such that

$$\left\| \sum_{k=1}^{\infty}((P^k F)(\theta', z) - \nabla f(\theta')) - \sum_{k=1}^{\infty}((P^k F)(\theta'', z) - \nabla f(\theta'')) \right\| \leq \tilde{C}_{2,Q}(1 + \|V\|^2)(\tilde{\varepsilon}_Q^n + n\|\theta' - \theta''\|) \tag{59}$$

for all $\theta', \theta'' \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$, $n \geq 0$ ($(P^0 F)(\theta, z)$ is defined as $F(\theta, z)$).

Let $\tilde{C}_Q = \max\{\tilde{C}_{1,Q}, \tilde{C}_{2,Q}\}$. Moreover, let $N_{Q,s}(t) = \lceil s \log t / \log \tilde{\varepsilon}_Q \rceil$ for $s, t \in (0, 1)$ and $N_{Q,s}(t) = 0$ for $s \in (0, 1)$, $t \in \{0\} \cup [1, \infty)$. Then, it can be concluded that there exists a real number $\tilde{K}_{Q,s} \in [1, \infty)$ such that

$$N_{Q,s}(t) + \tilde{\varepsilon}_Q^{N_{Q,s}(t)} \leq \tilde{K}_{Q,s} t^s \tag{60}$$

for all $t \in [0, \infty)$.

For $\theta \in \Theta$, $z = (x, y, u, V) \in \mathcal{S}_z$, let

$$\Phi(\theta, z) = \sum_{k=0}^{\infty}((P^k F)(\theta, z) - \nabla f(\theta)),$$

$$\varphi_{Q,s}(z) = \tilde{C}_Q \tilde{K}_{Q,s} \psi_Q(y)(1 + \|V\|^2).$$

Since

$$(P\varphi_{Q,s})(\theta, z) = \tilde{C}_Q \tilde{K}_{Q,s}(1 + \|H_\theta(u, V, y)\|^2) E(\psi_Q(Y_2)|X_1 = x) < \infty$$

for all $\theta \in \Theta$, $z = (x, y, u, V) \in \mathcal{S}_z$, we deduce from (58) that $\Phi(\cdot, \cdot)$ is well-defined, integrable and satisfies (54), (55) (notice that $(P\Phi)(\theta, z) = \sum_{k=1}^{\infty}((P^k F)(\theta, z) - \nabla f(\theta))$). On the other hand, (59), (60) imply

$$\|(P\Phi)(\theta', z) - (P\Phi)(\theta'', z)\| \leq \tilde{C}_Q \tilde{K}_{Q,s}(1 + \|V\|^2)\|\theta' - \theta''\|^s$$

for any $\theta', \theta'' \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$ (set $n = N_{Q,s}(\|\theta' - \theta''\|)$ in (59)). Thus, (56) is true for each $\theta', \theta'' \in Q$, $z = (x, y, u, V) \in \mathcal{S}_z$.

Let $\boldsymbol{\theta} = \{\theta_n\}_{n \geq 0}$ and $\boldsymbol{Y} = \{Y_n\}_{n \geq 1}$. Due to Lemma 3, we have

$$\begin{aligned}
\varphi_{Q,s}(Z_{n+1}) I_{\{\tau_Q > n\}} =& \tilde{C}_Q \tilde{K}_{Q,s} \psi_Q(Y_{n+1})(1 + \|H_{\boldsymbol{\theta},\boldsymbol{Y}}^{0:n}(U_0, V_0)\|^2) I_{\{\tau_Q > n\}} \\
\leq & 4\tilde{C}_Q \tilde{K}_{Q,s} C_{4,Q}^2 \psi_Q(Y_{n+1})(1 + \|V_0\|^2)
\end{aligned} \tag{61}$$

for each $n \geq 0$ (notice that $H^{0:n}_{\boldsymbol{\theta},\mathbf{Y}}(U_0, V_0)$ depends only on the first n elements of $\boldsymbol{\theta}$, and that $\theta_1, \dots, \theta_n \in Q$ is sufficient for (61) to hold). Consequently,

$$E\left(\varphi^2_{Q,s}(Z_{n+1})I_{\{\tau_Q > n\}} \middle| Z_1 = z\right) \leq 16\tilde{C}^2_Q \tilde{K}^2_{Q,s} C^4_{4,Q}(1 + \|V\|)^4 E(\psi^2_Q(Y_{n+1})|X_1 = x)$$
$$\leq 16\tilde{C}^2_Q \tilde{K}^2_{Q,s} C^4_{4,Q}(1 + \|V\|)^4 \max_{x' \in \mathcal{X}} \int \psi^2_Q(y')Q(dy'|x') < \infty$$

for all $z = (x, y, u, V) \in \mathcal{S}_z$, $n \geq 0$. Hence, (57) is true for all $z \in \mathcal{S}_z$. ■

Lemma 9: Suppose that Assumption 1 holds. Then, there exists a real number $s \in (0, 1)$ such that $\sum_{n=0}^{\infty} \alpha_n^{1+s}\gamma_n^r < \infty$.

Proof: Let $p = (2 + 2r)/(2 + r)$, $q = (2 + 2r)/r$, $s = (2 + r)/(2 + 2r)$. Then, using the Hölder inequality, we get

$$\sum_{n=0}^{\infty} \alpha_n^{1+s}\gamma_n^r = \sum_{n=1}^{\infty} (\alpha_n^2 \gamma_n^{2r})^{1/p} \left(\frac{\alpha_n}{\gamma_n^2}\right)^{1/q} \leq \left(\sum_{n=1}^{\infty} \alpha_n^2 \gamma_n^{2r}\right)^{1/p} \left(\sum_{n=1}^{\infty} \frac{\alpha_n}{\gamma_n^2}\right)^{1/q}.$$

Since $\gamma_{n+1}/\gamma_n = 1 + \alpha_n/\gamma_n = O(1)$ for $n \to \infty$ and

$$\sum_{n=1}^{\infty} \frac{\alpha_n}{\gamma_n^2} = \sum_{n=1}^{\infty} \frac{\gamma_{n+1} - \gamma_n}{\gamma_n^2} \leq \sum_{n=1}^{\infty} \left(\frac{\gamma_{n+1}}{\gamma_n}\right)^2 \int_{\gamma_n}^{\gamma_{n+1}} \frac{dt}{t^2} = \frac{1}{\gamma_1} \max_{n \geq 0} \left(\frac{\gamma_{n+1}}{\gamma_n}\right)^2,$$

it is obvious that $\sum_{n=0}^{\infty} \alpha_n^{1+s}\gamma_n^r$ converges. ■

Lemma 10: Suppose that Assumptions 1 – 4 hold. Then, there exists an event N_0 such that $P(N_0) = 0$ and such that $\sum_{n=0}^{\infty} \alpha_n \gamma_n^r \xi_n$, $\sum_{n=0}^{\infty} \alpha_n \xi_n$ and $\sum_{n=0}^{\infty} \phi_n$ converge on $\Lambda \setminus N_0$.

Proof: Let $Q \subset \Theta$ be an arbitrary compact set, while t is an arbitrary number from $[0, r]$. Moreover, let $\Psi : \Theta \to \mathbb{R}^{d_\theta \times d_\theta}$ be an arbitrary locally Lipschitz continuous function. Obviously, in order to prove the lemma, it is sufficient to demonstrate that $\sum_{n=0}^{\infty} \alpha_n \gamma_n^t \Psi(\theta_n)\xi_n$ and $\sum_{n=0}^{\infty} \phi''_n$ converge w.p.1 on $\bigcap_{n=0}^{\infty}\{\theta_n \in Q\}$ (to show the convergence of $\sum_{n=0}^{\infty} \alpha_n \gamma_n^r \xi_n$, set $t = r$ and $\Psi(\theta) = I$ for all $\theta \in \Theta$, where I stands for $d_\theta \times d_\theta$ unit matrix; to demonstrate the convergence of $\sum_{n=0}^{\infty} \phi'_n$, set $t = 0$ and $\Psi(\theta) = e(\nabla f(\theta))^T$ for each $\theta \in \Theta$, where $e = [1 \cdots 1]^T \in \mathbb{R}^{d_\theta}$).

Let $s \in (0, 1)$ be a real number such that $\sum_{n=0}^{\infty} \alpha_n^{1+s}\gamma_n^r < \infty$, while

$$\tilde{C}_Q = \max\left\{\|\nabla\Psi(\theta)\|, \frac{\|\Psi(\theta') - \Psi(\theta'')\|}{\|\theta' - \theta''\|^s}, \frac{\|\nabla f(\theta') - \nabla f(\theta'')\|}{\|\theta' - \theta''\|} : \theta, \theta', \theta'' \in Q\right\}.$$

Moreover, for $n \geq 1$, let

$$\psi_{1,n} = \Psi(\theta_n)(\Phi(\theta_n, Z_{n+1}) - (P\Phi)(\theta_n, Z_n)),$$
$$\psi_{2,n} = \Psi(\theta_n)((P\Phi)(\theta_n, Z_n) - (P\Phi)(\theta_{n-1}, Z_n)) + (\Psi(\theta_n) - \Psi(\theta_{n-1}))(P\Phi)(\theta_{n-1}, Z_n),$$
$$\psi_{3,n} = \Psi(\theta_n)(P\Phi)(\theta_n, Z_{n+1}).$$

Then, it is straightforward to verify

$$\sum_{i=1}^{n} \alpha_i \gamma_i^t \Psi(\theta_i)\xi_i = \sum_{i=1}^{n} \alpha_i \gamma_i^t \psi_{1,i} + \sum_{i=1}^{n} \alpha_i \gamma_i^t \psi_{2,i} + \sum_{i=0}^{n-1} (\alpha_{i+1}\gamma_{i+1}^t - \alpha_i \gamma_i^t)\psi_{3,i} - \alpha_n \gamma_n^t \psi_{3,n} + \alpha_0 \gamma_0^t \psi_{3,0} \tag{62}$$

for $n \geq 1$.

Owing to Assumption 1, we have

$$\alpha_n = \alpha_{n+1}(1+\alpha_n(\alpha_{n+1}^{-1}-\alpha_n^{-1})) = O(\alpha_{n+1}),$$
$$\alpha_n - \alpha_{n+1} = \alpha_n\alpha_{n+1}(\alpha_{n+1}^{-1}-\alpha_n^{-1}) = O(\alpha_{n+1}^2),$$
$$\gamma_{n+1}^t - \gamma_n^t = \gamma_n^t\left((1+\alpha_n/\gamma_n)^t - 1\right) = o(\alpha_n\gamma_n^t)$$

as $n\to\infty$. Consequently,

$$\sum_{n=0}^{\infty}\alpha_n^s\alpha_{n+1}\gamma_{n+1}^t = \sum_{n=0}^{\infty}(\alpha_n/\alpha_{n+1})^s\alpha_{n+1}^s\gamma_{n+1}^t < \infty, \tag{63}$$
$$\sum_{n=0}^{\infty}|\alpha_n\gamma_n^t - \alpha_{n+1}\gamma_{n+1}^t| \leq \sum_{n=0}^{\infty}\alpha_n|\gamma_n^t-\gamma_{n+1}^t| + \sum_{n=0}^{\infty}|\alpha_n-\alpha_{n+1}|\gamma_{n+1}^t < \infty. \tag{64}$$

On the other side, as a consequence of Lemma 8, we get

$$E_{\theta,z}\left(|\psi_{1,n}|^2 I_{\{\tau_Q>n\}}\right) \leq 2\tilde{C}_Q^2 E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right) + 2\tilde{C}_Q^2 E_{\theta,z}\left(\varphi_{Q,s}^2(Z_n)I_{\{\tau_Q>n-1\}}\right),$$
$$E_{\theta,z}\left(|\psi_{2,n}|I_{\{\tau_Q>n\}}\right) \leq 2\tilde{C}_Q E_{\theta,z}\left(\varphi_{Q,s}(Z_n)\|\theta_n-\theta_{n-1}\|^s I_{\{\tau_Q>n\}}\right) \leq 2\tilde{C}_Q\alpha_{n-1}^s E_{\theta,z}\left(\varphi_{Q,s}^2(Z_n)I_{\{\tau_Q>n-1\}}\right)$$

for all $\theta\in\Theta$, $z\in\mathcal{S}_z$, $n\geq 1$. Due to the same lemma, we have

$$E_{\theta,z}\left(|\psi_{3,n}|^2 I_{\{\tau_Q>n\}}\right) \leq \tilde{C}_Q^2 E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right),$$
$$E_{\theta,z}\left(|\phi_n''|I_{\{\tau_Q>n\}}\right) \leq \tilde{C}_Q E_{\theta,z}\left(\|\theta_{n+1}-\theta_n\|^2 I_{\{\tau_Q>n\}}\right) \leq \tilde{C}_Q\alpha_n^2 E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right)$$

for all $\theta\in\Theta$, $z\in\mathcal{S}_z$, $n\geq 1$. Then, Lemma 8 and (63) yield

$$E_{\theta,z}\left(\sum_{n=1}^{\infty}\alpha_n^2\gamma_n^{2t}|\psi_{1,n}|^2 I_{\{\tau_Q>n\}}\right) \leq 4\tilde{C}_Q^2\left(\sum_{n=1}^{\infty}\alpha_n^2\gamma_n^{2t}\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right) < \infty,$$
$$E_{\theta,z}\left(\sum_{n=1}^{\infty}\alpha_n\gamma_n^t|\psi_{2,n}|I_{\{\tau_Q>n\}}\right) \leq 2\tilde{C}_Q\left(\sum_{n=1}^{\infty}\alpha_{n-1}^s\alpha_n\gamma_n^t\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right) < \infty$$

for any $\theta\in\Theta$, $z\in\mathcal{S}_z$. On the other side, Lemma 8 and (64) imply

$$E_{\theta,z}\left(\sum_{n=1}^{\infty}|\alpha_n\gamma_n^t-\alpha_{n+1}\gamma_{n+1}^t|\,|\psi_{3,n}|I_{\{\tau_Q>n\}}\right)$$
$$\leq \tilde{C}_Q\left(\sum_{n=1}^{\infty}|\alpha_n\gamma_n^t-\alpha_{n+1}\gamma_{n+1}^t|\right)\sup_{n\geq 0}\left(E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right)\right)^{1/2} < \infty,$$
$$E_{\theta,z}\left(\sum_{n=1}^{\infty}\alpha_{n+1}^2\gamma_{n+1}^{2t}|\psi_{3,n}|^2 I_{\{\tau_Q>n\}}\right) \leq \tilde{C}_Q^2\left(\sum_{n=1}^{\infty}\alpha_{n+1}^2\gamma_{n+1}^{2t}\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right) < \infty,$$
$$E_{\theta,z}\left(\sum_{n=0}^{\infty}|\phi_n''|I_{\{\tau_Q>n\}}\right) \leq \tilde{C}_Q\left(\sum_{n=0}^{\infty}\alpha_n^2\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right) < \infty$$

for each $\theta\in\Theta$, $z\in\mathcal{S}_z$. Since

$$E_{\theta,z}\left(\psi_{1,n}I_{\{\tau_Q>n\}}|\mathcal{F}_n\right) = \Psi(\theta_n)\left(E_{\theta,z}\left(\Phi(\theta_n,Z_{n+1})|\mathcal{F}_n\right) - (P\Phi)(\theta_n,Z_n)\right)I_{\{\tau_Q>n\}} = 0$$

w.p.1 for every $\theta \in \Theta$, $z \in \mathcal{S}_z$, $n \geq 1$, it is clear that series

$$\sum_{n=1}^{\infty} \alpha_n \gamma_n^t \psi_{1,n}, \quad \sum_{n=1}^{\infty} \alpha_n \gamma_n^t \psi_{2,n}, \quad \sum_{n=1}^{\infty} (\alpha_n \gamma_n^t - \alpha_{n+1} \gamma_{n+1}^t) \psi_{3,n}, \quad \sum_{n=1}^{\infty} \phi_n''$$

converge w.p.1 on $\bigcap_{n=0}^{\infty} \{\theta_n \in Q\}$, and that $\lim_{n\to\infty} \alpha_n \gamma_n^t \psi_{3,n} = 0$ w.p.1 on the same event. Owing to this and (62), we have that $\sum_{n=0}^{\infty} \alpha_n \gamma_n^t \Psi(\theta_n) \xi_n$ is convergent w.p.1 on $\bigcap_{n=0}^{\infty} \{\theta_n \in Q\}$. ■

Lemma 11: Suppose that Assumption 1 – 4 hold. Then, on $\Lambda \setminus N_0$, $\lim_{n\to\infty} \nabla f(\theta_n) = 0$ and $\lim_{n\to\infty} f(\theta_n)$ exists.

Proof: Let $Q \subset \Theta$ be an arbitrary compact set, while ω is an arbitrary sample from $\bigcap_{n=0}^{\infty} \{\theta_n \in Q\} \setminus N_0$ (notice that all formulas which appear in the proof correspond to this ω). Obviously, in order to prove the lemma, it is sufficient to show that $\lim_{n\to\infty} f(\theta_n)$ exists and that $\lim_{n\to\infty} \nabla f(\theta_n) = 0$.

Since $\sum_{n=0}^{\infty} \phi_n$ converges and

$$\sum_{i=0}^{n-1} \alpha_i \|\nabla f(\theta_i)\|^2 = f(\theta_n) - f(\theta_0) - \sum_{i=0}^{n-1} \phi_i$$

for $n \geq 0$, we conclude $\sum_{n=0}^{\infty} \alpha_n \|\nabla f(\theta_n)\|^2 < \infty$ (also notice that $f(\cdot)$ is bounded on Q). As

$$f(\theta_n) = f(\theta_0) + \sum_{i=0}^{n-1} \alpha_i \|\nabla f(\theta_i)\|^2 + \sum_{i=0}^{n-1} \phi_i$$

for $n \geq 0$, it is clear that $\lim_{n\to\infty} f(\theta_n)$ exists.

Let $\tilde{C}_Q$ be a Lipschitz constant of $\nabla f(\cdot)$ on Q and an upper bound of $\|\nabla f(\cdot)\|$ on the same set. Now, we prove $\lim_{n\to\infty} \nabla f(\theta_n) = 0$. Suppose the opposite. Then, there exist $\varepsilon \in (0, \infty)$ and sequences $\{m_k\}_{k\geq 0}$, $\{n_k\}_{k\geq 0}$ (all depending on ω) such that $m_k < n_k < m_{k+1}$, $\|\nabla f(\theta_{m_k})\| \leq \varepsilon$, $\|\nabla f(\theta_{n_k})\| \geq 2\varepsilon$ for $k \geq 0$, and such that $\|\nabla f(\theta_n)\| \geq \varepsilon$ for $m_k < n \leq n_k$, $k \geq 0$. Therefore,

$$\varepsilon \leq \|\nabla f(\theta_{n_k}) - \nabla f(\theta_{m_k})\| \leq \tilde{C}_Q \|\theta_{n_k} - \theta_{m_k}\| \leq \tilde{C}_Q^2 \sum_{i=m_k}^{n_k-1} \alpha_i + \tilde{C}_Q \left\| \sum_{i=m_k}^{n_k-1} \alpha_i \xi_i \right\| \tag{65}$$

for $k \geq 0$. We also have

$$\varepsilon^2 \sum_{i=m_k+1}^{n_k} \alpha_i \leq \sum_{i=m_k+1}^{\infty} \alpha_i \|\nabla f(\theta_i)\|^2$$

for $k \geq 0$. Consequently, $\lim_{k\to\infty} \sum_{i=m_k}^{n_k-1} \alpha_i = 0$. However, this is not possible, since the limit process $k \to \infty$ applied to (65) would imply

$$\varepsilon \leq \lim_{k\to\infty} \|\nabla f(\theta_{n_k}) - \nabla f(\theta_{m_k})\| = 0.$$

Hence, $\lim_{n\to\infty} \nabla f(\theta_n) = 0$. ■

D. Convergence and Convergence Rate

In this subsection, using the results of Subsections IV-B, IV-C (Corollary 5, Lemmas 9, 10), the convergence and convergence rate of recursion (1) – (3) are analyzed and Theorems 2 and 3 are proved.

Throughout the subsection, we use the following notation. For $t \in (0, \infty)$, $n \geq 0$, let

$$a(n,t) = \max\{k \geq n : \gamma_k - \gamma_n \leq t\}.$$

For $0 \leq n \leq k$, let

$$\begin{aligned}
\zeta_n &= \sup_{k \geq n} \left\| \sum_{i=n}^{k} \alpha_i \xi_i \right\|, \\
\varepsilon'_{n,k} &= \sum_{i=n}^{k-1} \alpha_i \xi_i, \\
\varepsilon''_{n,k} &= \sum_{i=n}^{k-1} \alpha_i (\nabla f(\theta_i) - \nabla f(\theta_n)), \\
\phi'_{n,k} &= (\nabla f(\theta_n))^T (\varepsilon'_{n,k} + \varepsilon''_{n,k}), \\
\phi''_{n,k} &= \int_0^1 (\nabla f(\theta_n + t(\theta_k - \theta_n)) - \nabla f(\theta_n))^T (\theta_k - \theta_n) dt,
\end{aligned}$$

while $\varepsilon_{n,k} = \varepsilon'_{n,k} + \varepsilon''_{n,k}$ and $\phi_{n,k} = \phi'_{n,k} + \phi''_{n,k}$. Then, it is straightforward to verify

$$\begin{aligned}
\theta_k =& \theta_n + \sum_{i=n}^{k-1} \alpha_i \nabla f(\theta_i) + \varepsilon'_{n,k} \\
=& \theta_n + (\gamma_k - \gamma_n) \nabla f(\theta_n) + \varepsilon_{n,k}, \qquad (66) \\
f(\theta_k) =& f(\theta_n) + (\gamma_k - \gamma_n) \|\nabla f(\theta_n)\|^2 + \phi_{n,k} \qquad (67)
\end{aligned}$$

for $0 \leq n \leq k$.

Besides the notation introduced in the previous paragraph, we also rely on the following notation in this subsection. For a compact set $Q \subset \Theta$, $C_Q \in [1, \infty)$ denotes an upper bound of $\|\nabla f(\cdot)\|$ on Q and a Lipschitz constant of $\nabla f(\cdot)$ on the same set. $\hat{A}$ is the set of the accumulation points of $\{\theta_n\}_{n \geq 0}$, while

$$\hat{f} = \liminf_{n \to \infty} f(\theta_n).$$

$\hat{\rho}$ and $\hat{B}$, $\hat{Q}$ are a random quantity and random sets (respectively) defined by

$$\hat{\rho} = d(\hat{A}, \partial\Theta)/2, \qquad \hat{B} = \bigcup_{\theta \in \hat{A}} \left\{ \theta' \in \mathbb{R}^{d_\theta} : \|\theta' - \theta\| \leq \min\{\delta_\theta, \hat{\rho}\} \right\}, \qquad \hat{Q} = \text{cl}(\hat{B})$$

on Λ, and by

$$\hat{\rho} = 0, \qquad \hat{B} = \hat{A}, \qquad \hat{Q} = \hat{A}$$

otherwise. Overriding the definition of $\hat{\mu}$ in Theorem 3, we specify random quantities $\hat{\delta}$, $\hat{\mu}$, $\hat{C}$, $\hat{C}$ as

$$\hat{\delta} = \delta_{\hat{Q}, \hat{f}}, \qquad \hat{\mu} = \mu_{\hat{Q}, \hat{f}}, \qquad \hat{C} = C_{\hat{Q}, \hat{f}}, \qquad \hat{M} = M_{\hat{Q}, \hat{f}} \qquad (68)$$

on Λ, and as

$$\hat{\delta} = 1, \qquad \hat{\mu} = 2, \qquad \hat{C} = 1, \qquad \hat{M} = 1$$

otherwise ($\delta_{Q,a}$, $\mu_{Q,a}$, $M_{Q,a}$ are introduced in the statement of Corollary 5; later, once Theorem 2 is proved, it will be clear that the definitions of $\hat{\mu}$ provided in Theorem 3 and in (68) are equivalent). Random quantities $\hat{p}$, $\hat{q}$, $\hat{r}$ are defined in the same way as in (5). Functions $u(\cdot)$ and $v(\cdot)$ are defined by

$$u(\theta) = \hat{f} - f(\theta), \qquad v(\theta) = \begin{cases} (1/u(\theta))^{1/\hat{p}}, & \text{if } u(\theta) > 0 \\ 0, & \text{otherwise} \end{cases}$$

for $\theta \in \Theta$.

Obviously, on event Λ, $\hat{Q}$ is compact and satisfies $\hat{A} \subset \text{int}\hat{Q}$, $\hat{Q} \subset \Theta$. Thus, $\hat{\mu}$, $\hat{M}$, $\hat{p}$, $\hat{q}$, $\hat{r}$, $v(\cdot)$ are are well-defined on the same event (what happens with these quantities outside Λ does not affect the results provided in this subsection). On the other side, Corollary 5 implies

$$|f(\theta) - \hat{f}| \leq \hat{M} \|\nabla f(\theta)\|^{\hat{\mu}} \tag{69}$$

on Λ for all $\theta \in \hat{Q}$ satisfying $|f(\theta) - \hat{f}| \leq \hat{\delta}$.

Lemma 12: Suppose that Assumptions 1 – 4 hold. Then, $\lim_{n\to\infty} \gamma_n^r \zeta_n = 0$ on $\Lambda \setminus N_0$ (N_0 is specified in the statement of Lemma 10).

Proof: It is straightforward to verify

$$\sum_{i=n}^{k} \gamma_i \xi_i = \gamma_{k+1}^{-r} \sum_{j=n}^{k} \alpha_j \gamma_j^r \xi_j + \sum_{i=n}^{k} (\gamma_i^{-r} - \gamma_{i+1}^{-r}) \sum_{j=n}^{i} \alpha_j \gamma_j^r \xi_j$$

for $0 \leq n \leq k$. Therefore,

$$\left\| \sum_{i=n}^{k} \gamma_i \xi_i \right\| \leq \left(\gamma_{k+1}^{-r} + \sum_{i=n}^{k} (\gamma_i^{-r} - \gamma_{i+1}^{-r}) \right) \sup_{i \geq n} \left\| \sum_{j=n}^{i} \alpha_j \gamma_j^r \xi_j \right\| = \gamma_n^{-r} \sup_{i \geq n} \left\| \sum_{j=n}^{i} \alpha_j \gamma_j^r \xi_j \right\|$$

for $0 \leq n \leq k$. Consequently, Lemma 10 implies

$$\limsup_{n\to\infty} \gamma_n^r \zeta_n = \limsup_{n\to\infty} \sup_{k \geq n} \left\| \sum_{i=n}^{k} \alpha_i \gamma_i^r \xi_i \right\| = 0$$

on $\Lambda \setminus N_0$. ■

Lemma 13: Suppose that Assumptions 1 – 4 hold. Let $\hat{C}_1 = (16\hat{p}\hat{M})^{2\hat{p}}$ (notice that $1 \leq \hat{C}_1 < \infty$ everywhere). Then, there exist a random quantity $\hat{t}$ and an integer-valued random variable σ such that $0 < \hat{t} < 1$, $0 \leq \sigma < \infty$ everywhere and such that

$$\max_{n \leq k \leq a(n,\hat{t})} \|\varepsilon_{n,k}\| \leq (\hat{t}/\hat{C}_1)(\gamma_n^{-r} + \|\nabla f(\theta_n)\|), \tag{70}$$

$$\max_{n \leq k \leq a(n,\hat{t})} |\phi_{n,k}| \leq (\hat{t}/\hat{C}_1)(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2), \tag{71}$$

$$f(\theta_n) - f(\theta_{a(n,\hat{t})}) + 2^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 \leq (\hat{t}/\hat{C}_1)\gamma_n^{-2r}, \tag{72}$$

$$f(\theta_n) - f(\theta_{a(n,\hat{t})}) + 2^{-1}\|\nabla f(\theta_n)\| \|\theta_{a(n,\hat{t})} - \theta_n\| \leq (\hat{t}/\hat{C}_1)\gamma_n^{-2r} \tag{73}$$

on $\Lambda \setminus N_0$ for $n > \sigma$.

Proof: Let $\tilde{C}_1 = 2\hat{C}\exp(\hat{C})$, $\tilde{C}_2 = 2\hat{C}\tilde{C}_1$, $\tilde{C}_3 = 2\hat{C}\tilde{C}_2^2 + \tilde{C}_2$ and $\tilde{C}_4 = \tilde{C}_2 + \tilde{C}_3$, while $\hat{t} = 1/(2\hat{C}_1\tilde{C}_4)$. Moreover, let

$$\begin{aligned}\tilde{\sigma}_1 =& \max\left(\left\{n \geq 0 : \theta_n \notin \hat{Q}\right\} \cup \{0\}\right),\\ \tilde{\sigma}_2 =& \max\left(\left\{n \geq 0 : \alpha_n > \hat{t}/4\right\} \cup \{0\}\right),\\ \tilde{\sigma}_3 =& \max\left(\left\{n \geq 0 : \gamma_n^r \zeta_n > \hat{t}/(2\hat{C}_1\tilde{C}_4)\right\} \cup \{0\}\right)\end{aligned}$$

while $\sigma = \max\{\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3\} I_{\Lambda \setminus N_0}$. Then, it is obvious that σ is well-defined, while Lemma 12 implies $0 \leq \sigma < \infty$ everywhere. We also have

$$\max\{\tilde{C}_2\gamma_n^r\zeta_n, \tilde{C}_3\gamma_n^r\zeta_n, \tilde{C}_3\gamma_n^{2r}\zeta_n^2, \tilde{C}_4\gamma_n^r\zeta_n, \tilde{C}_4\gamma_n^{2r}\zeta_n^2\} \leq 2^{-1}\hat{C}_1^{-1}\hat{t}, \tag{74}$$

$$\max\{\tilde{C}_2\hat{t}^2, \tilde{C}_3\hat{t}^2, \tilde{C}_4\hat{t}^2\} \leq 2^{-1}\hat{C}_1^{-1}\hat{t}, \tag{75}$$

$$\hat{t} \geq \gamma_{a(n,\hat{t})} - \gamma_n = \gamma_{a(n,\hat{t})+1} - \gamma_n - \alpha_{a(n,\hat{t})} \geq 3\hat{t}/4 \tag{76}$$

on $\Lambda \setminus N_0$ for $n > \sigma$.

Let ω be an arbitrary sample from Λ (notice that all formulas which follow in the proof correspond to this ω). Since $\theta_n \in \hat{Q}$ for $n > \sigma$, we have

$$\begin{aligned}\|\nabla f(\theta_k)\| \leq& \|\nabla f(\theta_n)\| + \|\nabla f(\theta_k) - \nabla f(\theta_n)\|\\ \leq& \|\nabla f(\theta_n)\| + \hat{C}\|\theta_k - \theta_n\|\\ \leq& \|\nabla f(\theta_n)\| + \hat{C}\sum_{i=n}^{k-1}\alpha_i\|\nabla f(\theta_i)\| + \hat{C}\|\varepsilon'_{n,k}\|\\ \leq& \hat{C}(\zeta_n + \|\nabla f(\theta_n)\|) + \hat{C}\sum_{i=n}^{k-1}\alpha_i\|\nabla f(\theta_i)\|\end{aligned}$$

for $\sigma < n \leq k$. Then, Bellman-Gronwall inequality yields

$$\|\nabla f(\theta_k)\| \leq \hat{C}(\zeta_n + \|\nabla f(\theta_n)\|)\exp(\hat{C}(\gamma_k - \gamma_n)) \leq \hat{C}\exp(\hat{C})(\zeta_n + \|\nabla f(\theta_n)\|)$$

for $\sigma < n \leq k \leq a(n,1)$. Consequently,

$$\begin{aligned}\|\theta_k - \theta_n\| \leq& \sum_{i=n}^{k-1}\alpha_i\|\nabla f(\theta_i)\| + \|\varepsilon'_{n,k}\|\\ \leq& \zeta_n + \hat{C}\exp(\hat{C})(\zeta_n + \|\nabla f(\theta_n)\|)(\gamma_k - \gamma_n)\\ \leq& \tilde{C}_1(\zeta_n + (\gamma_k - \gamma_n)\|\nabla f(\theta_n)\|)\end{aligned}$$

for $\sigma < n \leq k \leq a(n,1)$. Therefore,

$$\begin{aligned}\|\varepsilon_{n,k}\| \leq& \|\varepsilon'_{n,k}\| + \hat{C}\sum_{i=n}^{k-1}\alpha_i\|\theta_i - \theta_n\|\\ \leq& \zeta_n + \hat{C}\tilde{C}_1((\gamma_k - \gamma_n)\zeta_n + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|)\\ \leq& \tilde{C}_2(\zeta_n + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|)\end{aligned} \tag{77}$$

for $\sigma < n \leq k \leq a(n,1)$ (notice that $\gamma_k - \gamma_n \leq 1$ for $n \leq k \leq a(n,1)$). Thus,

$$\begin{aligned}
\|\phi_{n,k}\| \leq & \|\nabla f(\theta_n)\| \|\varepsilon_{n,k}\| + \hat{C}\|\theta_k - \theta_n\|^2 \\
\leq & \tilde{C}_2(\zeta_n\|\nabla f(\theta_n)\| + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|^2) + 2\hat{C}\tilde{C}_1^2(\zeta_n^2 + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|^2) \\
\leq & \tilde{C}_3(\zeta_n^2 + \zeta_n\|\nabla f(\theta_n)\| + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|^2)
\end{aligned} \tag{78}$$

for $\sigma < n \leq k \leq a(n,1)$. On the other side, combining (66), (67), we get

$$\begin{aligned}
f(\theta_k) - f(\theta_n) = & \|\nabla f(\theta_n)\| \|(\gamma_k - \gamma_n)\nabla f(\theta_n)\| + \phi_{n,k} \\
= & \|\nabla f(\theta_n)\| \|\theta_k - \theta_n + \varepsilon_{n,k}\| + \phi_{n,k} \\
\geq & \|\nabla f(\theta_n)\|(\|\theta_k - \theta_n\| - \|\varepsilon_{n,k}\|) - |\phi_{n,k}|
\end{aligned}$$

for $0 \leq n \leq k$. Then, (77), (78) yield

$$\begin{aligned}
f(\theta_n) - f(\theta_k) + \|\nabla f(\theta_n)\| \|\theta_k - \theta_n\| \leq & \|\nabla f(\theta_n)\| \|\varepsilon_{n,k}\| + |\phi_{n,k}| \\
\leq & \tilde{C}_3\zeta_n^2 + (\tilde{C}_2 + \tilde{C}_3)(\zeta_n\|\nabla f(\theta_n)\| + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|^2) \\
\leq & \tilde{C}_4(\zeta_n^2 + \zeta_n\|\nabla f(\theta_n)\| + (\gamma_k - \gamma_n)^2\|\nabla f(\theta_n)\|^2)
\end{aligned} \tag{79}$$

for $\sigma < n \leq k \leq a(n,1)$.

Owing to (74), (75), (77), (78), we have

$$\begin{aligned}
\|\varepsilon_{n,k}\| \leq & \tilde{C}_2\zeta_n + \tilde{C}_2\hat{t}^2\|\nabla f(\theta_n)\| \\
\leq & \hat{C}_1^{-1}\hat{t}(\gamma_n^{-r} + \|\nabla f(\theta_n)\|),
\end{aligned} \tag{80}$$

$$\begin{aligned}
|\phi_{n,k}| \leq & \tilde{C}_3\zeta_n^2 + \tilde{C}_3\zeta_n\|\nabla f(\theta_n)\| + \tilde{C}_3\hat{t}^2\|\nabla f(\theta_n)\|^2 \\
\leq & 2^{-1}\hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \gamma_n^{-r}\|\nabla f(\theta_n)\| + \|\nabla f(\theta_n)\|^2) \\
\leq & \hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2)
\end{aligned} \tag{81}$$

for $\sigma < n \leq k \leq a(n,\hat{t})$ (notice that $\gamma_k - \gamma_n \leq \hat{t}$ for $n \leq k \leq a(n,\hat{t})$). Due to (67), (76), (81), we have also

$$\begin{aligned}
f(\theta_n) - f(\theta_{a(n,\hat{t})}) \leq & -(\gamma_{a(n,\hat{t})} - \gamma_n)\|\nabla f(\theta_n)\|^2 + |\phi_{n,a(n,\hat{t})}| \\
\leq & -(3\hat{t}/4)\|\nabla f(\theta_n)\|^2 + \hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2) \\
= & -(3/4 - \hat{C}_1^{-1})\hat{t}\|\nabla f(\theta_n)\|^2 + \hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} \\
\leq & -2^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 + \hat{C}_1^{-1}\hat{t}\gamma_n^{-2r}
\end{aligned} \tag{82}$$

for $n > \sigma$ (notice that $\hat{C}_1 \geq 4$). Consequently,

$$\hat{C}_1^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 \leq 2^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 \leq \hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} + (f(\theta_{a(n,\hat{t})}) - f(\theta_n)) \tag{83}$$

for $n > \sigma$. On the other side, (74) – (76), (79), (83) imply

$$
\begin{aligned}
f(\theta_n) - f(\theta_{a(n,\hat{t})}) + \|\nabla f(\theta_n)\| \|\theta_{a(n,\hat{t})} - \theta_n\| \leq & \tilde{C}_4(\zeta_n^2 + \zeta_n \|\nabla f(\theta_n)\| + \hat{t}^2 \|\nabla f(\theta_n)\|^2) \\
\leq & 2^{-1}\hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \gamma_n^{-r}\|\nabla f(\theta_n)\| + \|\nabla f(\theta_n)\|^2) \\
\leq & \hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2) \\
\leq & 2\hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} + (f(\theta_{a(n,\hat{t})}) - f(\theta_n))
\end{aligned}
$$

for $n > \sigma$. Therefore,

$$
2(f(\theta_n) - f(\theta_{a(n,\hat{t})})) + \|\nabla f(\theta_n)\| \|\theta_{a(n,\hat{t})} - \theta_n\| \leq 2\hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} \tag{84}
$$

for $n > \sigma$. Then, (70) – (73) directly follow from (80), (81), (82), (84). ■

Lemma 14: Suppose that Assumptions 1 – 4 hold. Let $\hat{C}_2 = 4\hat{p}\hat{M}^2$ (notice that $1 \leq \hat{C}_2 < \infty$ everywhere). Then, there exists an integer-valued random variable τ such that $0 \leq \tau < \infty$ everywhere and such that

$$
\left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/4)\|\nabla f(\theta_n)\|^2 \right) I_{A_n} \leq 0, \tag{85}
$$

$$
\left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/\hat{C}_2) u(\theta_n) \right) I_{B_n} \leq 0, \tag{86}
$$

$$
\left(v(\theta_{a(n,\hat{t})}) - v(\theta_n) - \hat{t}/\hat{C}_2 \right) I_{C_n} \geq 0 \tag{87}
$$

on $\Lambda \setminus N_0$ for $n > \tau$, where

$$
\begin{aligned}
A_n = & \{\gamma_n^{\hat{p}} |u(\theta_n)| \geq 1\} \cup \{\gamma_n^r \|\nabla f(\theta_n)\| \geq 1\}, \\
B_n = & \{\gamma_n^{\hat{p}} u(\theta_n) \geq 1\} \cap \{\hat{\mu} = 2\}, \\
C_n = & \{\gamma_n^{\hat{p}} u(\theta_n) \geq 1\} \cap \{u(\theta_{a(n,\hat{t})}) > 0\} \cap \{\hat{\mu} < 2\}
\end{aligned}
$$

($\hat{t}$ is specified in the statement of Lemma 13).

Remark: Inequalities (85) – (87) can be interpreted in the following way: Relations

$$
\left(\gamma_n^{\hat{p}} |u(\theta_n)| \geq 1 \ \vee \ \gamma_n^r \|\nabla f(\theta_n)\| \geq 1 \right) \ \wedge \ n > \tau \implies u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/4)\|\nabla f(\theta_n)\|^2, \tag{88}
$$

$$
\gamma_n^{\hat{p}} u(\theta_n)| \geq 1 \ \wedge \ \hat{\mu} = 2 \ \wedge \ n > \tau \implies u(\theta_{a(n,\hat{t})}) \leq (1 - \hat{t}/\hat{C}_2) u(\theta_n), \tag{89}
$$

$$
\gamma_n^{\hat{p}} u(\theta_n) \geq 1 \ \wedge \ \hat{\mu} < 2 \ \wedge \ n > \tau \implies v(\theta_{a(n,\hat{t})}) - v(\theta_n) \geq \hat{t}/\hat{C}_2 \tag{90}
$$

are true on $\Lambda \setminus N_0$.

Proof: Let

$$
\begin{aligned}
\tilde{\tau}_1 = & \max\left(\left\{ n \geq 0 : \theta_n \notin \hat{Q} \right\} \cup \{0\} \right), \\
\tilde{\tau}_2 = & \max\left(\left\{ n \geq 0 : |u(\theta_n)| > \hat{\delta} \right\} \cup \{0\} \right)
\end{aligned}
$$

and $\tau = \max\{\sigma, \tilde{\tau}_1, \tilde{\tau}_2\} I_{\Lambda \setminus N_0}$. Then, it is obvious that τ is well-defined, while Lemma 11 implies $0 \leq \tau < \infty$ everywhere. On the other side, since $\tau \geq \sigma$ on $\Lambda \setminus N_0$, Lemma 13 (inequality (72)) implies

$$
u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/2)\|\nabla f(\theta_n)\|^2 + (\hat{t}/\hat{C}_1)\gamma_n^{-2r} \tag{91}
$$

on $\Lambda \setminus N_0$ for $n > \tau$. As $\theta_n \in \hat{Q}$, $|u(\theta_n| \leq \hat{\delta}$ on $\Lambda \setminus N_0$ for $n > \tau$, (69) (i.e., Corollary 5) yields

$$|u(\theta_n)| \leq \hat{M} \|\nabla f(\theta_n)\|^{\hat{\mu}} \tag{92}$$

on $\Lambda \setminus N_0$ for $n > \tau$.

Let ω be an arbitrary sample from $\Lambda \setminus N_0$ (notice that all formulas which follow in the proof correspond to this ω). First, we show (85). We proceed by contradiction: Suppose that (85) is violated for some $n > \tau$. Consequently,

$$u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/4)\|\nabla f(\theta_n)\|^2 > 0 \tag{93}$$

and at least one of the following two inequalities is true:

$$|u(\theta_n)| \geq \gamma_n^{-\hat{p}}, \qquad \|\nabla f(\theta_n)\| \geq \gamma_n^{-r}. \tag{94}$$

If $|u(\theta_n)| \geq \gamma_n^{-\hat{p}}$, then (92) implies

$$\|\nabla f(\theta_n)\|^2 \geq \left(|u(\theta_n)|/\hat{M}\right)^{2/\hat{\mu}} \geq (1/\hat{M})^{2/\hat{\mu}} \gamma_n^{-2\hat{p}/\hat{\mu}} \geq (4/\hat{C}_1)\gamma_n^{-2r}$$

(notice that $\hat{p}/\hat{\mu} \leq r$, $4\hat{M}^{2/\hat{\mu}} \leq 4\hat{M}^2 \leq \hat{C}_1$). Thus, as a result of one of (94), we get

$$\|\nabla f(\theta_n)\|^2 \geq (4/\hat{C}_1)\gamma_n^{-2r},$$

i.e., $(\hat{t}/4)\|\nabla f(\theta_n)\|^2 \geq (\hat{t}/\hat{C}_1)\gamma_n^{-2r}$. Then, (91) implies

$$u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/4)\|\nabla f(\theta_n)\|^2, \tag{95}$$

which directly contradicts (93). Hence, (85) is true for $n > \tau$. Owing to this, (92) and the fact that $B_n \subset A_n$ for $n \geq 0$, we obtain

$$\begin{aligned}\left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/\hat{C}_2)u(\theta_n)\right) I_{B_n} &\leq \left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{M}\hat{t}/\hat{C}_2)\|\nabla f(\theta_n)\|^2\right) I_{B_n} \\ &\leq \left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/4)\|\nabla f(\theta_n)\|^2\right) I_{B_n} \leq 0\end{aligned}$$

for $n > \tau$ (notice that $u(\theta_n) > 0$ on B_n; also notice that $4\hat{M} \leq \hat{C}_2$). Thus, (86) is satisfied.

Now, let us prove (87). To do so, we again use contradiction: Suppose that (87) does not hold for some $n > \tau$. Consequently, we have $\hat{\mu} < 2$, $u(\theta_{a(n,\hat{t})}) > 0$ and

$$\gamma_n^{\hat{p}} u(\theta_n) \geq 1, \tag{96}$$

$$v(\theta_{a(n,\hat{t})}) - v(\theta_n) < \hat{t}/\hat{C}_2. \tag{97}$$

Combining (96) with (already proved) (85), we get (95). On the other side, (92) yields

$$\|\nabla f(\theta_n)\|^2 \geq \left(u(\theta_n)/\hat{M}\right)^{2/\hat{\mu}} \geq \hat{M}^{-2}(u(\theta_n))^{1+1/\hat{p}}$$

(notice that $0 < u(\theta_n) \leq \hat{\delta} \leq 1$, $2/\hat{\mu} = 1 + 1/(\hat{\mu}\hat{r}) \leq 1 + 1/\hat{p}$). Therefore, (95) implies

$$\begin{aligned}\frac{\hat{t}}{4} \leq \frac{u(\theta_n) - u(\theta_{a(n,\hat{t})})}{\|\nabla f(\theta_n)\|^2} \leq & \hat{M}^2 \frac{u(\theta_n) - u(\theta_{a(n,\hat{t})})}{(u(\theta_n))^{1+1/\hat{p}}} \\ = & \hat{M}^2 \int_{u(\theta_{a(n,\hat{t})})}^{u(\theta_n)} \frac{du}{(u(\theta_n))^{1+1/\hat{p}}} \\ \leq & \hat{M}^2 \int_{u(\theta_{a(n,\hat{t})})}^{u(\theta_n)} \frac{du}{u^{1+1/\hat{p}}} \\ = & \frac{\hat{C}_2}{4}(v(\theta_{a(n,\hat{t})}) - v(\theta_n)).\end{aligned}$$

Thus, $v(\theta_{a(n,\hat{t})}) - v(\theta_n) \geq \hat{t}/\hat{C}_2$, which directly contradicts (97). Hence, (86) is satisfied for $n > \tau$. ■

Lemma 15: Suppose that Assumptions 1 – 4 hold. Then,

$$\gamma_n^{\hat{p}} u(\theta_n) \geq -1, \tag{98}$$

$$\|\nabla f(\theta_n)\|^2 \leq (4/\hat{t}) \left(\varphi(u(\theta_n)) + \gamma_n^{-\hat{p}}\right) \tag{99}$$

on $\Lambda \setminus N_0$ for $n > \tau$, where function $\varphi(\cdot)$ is defined by $\varphi(x) = x\, \mathrm{I}_{(0,\infty)}(x)$, $x \in \mathbb{R}$.

Proof: Let ω be an arbitrary sample from $\Lambda \setminus N_0$ (notice that all formulas that follow in the proof correspond to this ω). First, we prove (98). To do so, we use contradiction: Assume that (98) is not satisfied for some $n_0 > \tau$, and define recursively $n_{k+1} = a(n_k, \hat{t})$ for $k \geq 0$. Now, let us show by induction that $\{u(\theta_{n_k})\}_{k \geq 0}$ is non-increasing: Suppose that $u(\theta_{n_l}) \leq u(\theta_{n_{l-1}})$ for $0 \leq l \leq k$ and some $k \geq 1$. Consequently,

$$u(\theta_{n_k}) \leq u(\theta_{n_0}) \leq -\gamma_{n_0}^{-\hat{p}} \leq -\gamma_{n_k}^{-\hat{p}}.$$

Then, Lemma 14 (relations (85), (88)) yields

$$u(\theta_{n_{k+1}}) - u(\theta_{n_k}) \leq -(\hat{t}/4)\|\nabla f(\theta_{n_k})\|^2 \leq 0,$$

i.e., $u(\theta_{n_{k+1}}) \leq u(\theta_{n_k})$. Thus, $\{u(\theta_{n_k})\}_{k \geq 0}$ is non-increasing. Therefore,

$$\limsup_{n \to \infty} u(\theta_{n_k}) \leq u(\theta_{n_0}) < 0.$$

However, this is not possible, as $\lim_{n \to \infty} u(\theta_n) = 0$ (due to Lemma 11). Hence, (98) indeed holds for $n > \tau$.

Now, (99) is demonstrated. Again, we proceed by contradiction: Suppose that (99) is violated for some $n > \tau$. Consequently,

$$\|\nabla f(\theta_n)\|^2 \geq (4/\hat{t})\gamma_n^{-\hat{p}} \geq \gamma_n^{-2r}$$

(notice that $\hat{p} \leq \hat{\mu} r \leq 2r$), which, together with Lemma 14 (relations (85), (88)), yields

$$u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/4)\|\nabla f(\theta_n)\|^2.$$

Then, (98) implies

$$\|\nabla f(\theta_n)\|^2 \leq (4/\hat{t}) \left(u(\theta_n) - u(\theta_{a(n,\hat{t})}) \right) \leq (4/\hat{t}) \left(\varphi(u(\theta_n)) + \gamma_n^{-\hat{p}}\right).$$

However, this directly contradicts our assumption that n violates (99). Thus, (99) is satisfied for $n > \tau$. ■

Lemma 16: Suppose that Assumptions 1 – 4 hold. Let $\hat{C}_3 = 2\hat{C}_2^{\hat{p}}$. Then,

$$\liminf_{n\to\infty} \gamma_n^{\hat{p}} u(\theta_n) \leq \hat{C}_3 \tag{100}$$

on $\Lambda \setminus N_0$.

Proof: We prove the lemma by contradiction: Assume that (100) is violated for some sample ω from $\Lambda \setminus N_0$ (notice that the formulas which follow in the proof correspond to this ω). Consequently, there exists $n_0 > \tau$ such that

$$\gamma_n^{\hat{p}} u(\theta_n) \geq \hat{C}_3 \tag{101}$$

for $n \leq n_0$.

Let $\{n_k\}_{k\geq 0}$ be defined recursively as $n_{k+1} = a(n_k, \hat{t})$ for $k \geq 0$. In what follows in the proof, we consider separately the cases $\hat{\mu} < 2$ and $\hat{\mu} = 2$.

Case $\hat{\mu} < 2$: Owing to Lemma 14 (relations (87), (90)) and (101), we have

$$v(\theta_{n_{k+1}}) - v(\theta_{n_k}) \geq \hat{t}/\hat{C}_2 \geq (\gamma_{n_{k+1}} - \gamma_{n_k})/\hat{C}_2$$

for $k \geq 0$ (notice that $\gamma_n^{\hat{p}} u(\theta_n) \geq 1$ due to (101); also notice that $\gamma_{n_{k+1}} - \gamma_{n_k} \leq \hat{t}$). Therefore,

$$v(\theta_{n_k}) \geq v(\theta_{n_0}) + (1/\hat{C}_2) \sum_{i=0}^{k-1} (\gamma_{n_{i+1}} - \gamma_{n_i}) = v(\theta_{n_0}) + (\gamma_{n_k} - \gamma_{n_0})/\hat{C}_2$$

for $k \geq 0$. Then, (101) implies

$$\left(v(\theta_{n_0})/\gamma_{n_k} + (1 - \gamma_{n_0}/\gamma_{n_k})/\hat{C}_2 \right)^{-\hat{p}} \geq (v(\theta_{n_k})/\gamma_{n_k})^{-\hat{p}} = \gamma_{n_k}^{\hat{p}} u(\theta_{n_k}) \geq \hat{C}_3$$

for $k \geq 0$. However, this is impossible, since the limit process $k \to \infty$ (applied to the previous relation) yields $\hat{C}_3 \leq \hat{C}_2^{\hat{p}}$. Hence, (100) holds when $\hat{\mu} < 2$.

Case $\hat{\mu} = 2$: Due to Lemma 14 (relations (86), (89)) and (101), we have

$$u(\theta_{n_{k+1}}) \leq (1 - \hat{t}/\hat{C}_2) u(\theta_{n_k}) \leq \left(1 - (\gamma_{n_{k+1}} - \gamma_{n_k})/\hat{C}_2 \right) u(\theta_{n_k})$$

for $k \geq 0$. Consequently,

$$\begin{aligned} u(\theta_{n_k}) \leq & u(\theta_{n_0}) \prod_{i=0}^{k-1} \left(1 - (\gamma_{n_{i+1}} - \gamma_{n_i})/\hat{C}_2 \right) \\ \leq & u(\theta_{n_0}) \exp\left(-(1/\hat{C}_2) \sum_{i=0}^{k-1} (\gamma_{n_{i+1}} - \gamma_{n_i}) \right) \\ = & u(\theta_{n_0}) \exp\left(-(\gamma_{n_k} - \gamma_{n_0})/\hat{C}_2 \right) \end{aligned}$$

for $k \geq 0$. Then, (101) yields

$$u(\theta_{n_0}) \gamma_{n_k}^{\hat{p}} \exp\left(-(\gamma_{n_k} - \gamma_{n_0})/\hat{C}_2 \right) \geq \gamma_{n_k}^{\hat{p}} u(\theta_{n_k}) \geq \hat{C}_3$$

for $k \geq 0$. However, this is not possible, as the limit process $k \to \infty$ (applied to the previous relation) implies $\hat{C}_3 \leq 0$. Thus, (100) holds in the case $\hat{\mu} = 2$, too. ■

Lemma 17: Suppose that Assumptions 1 – 4 hold. Let $\hat{C}_4 = 6\hat{C}_3$. Then,

$$\limsup_{n\to\infty} \gamma_n^{\hat{p}} u(\theta_n) \leq \hat{C}_4 \tag{102}$$

on $\Lambda \setminus N_0$.

Proof: We use contradiction to prove the lemma: Suppose that (102) is violated for some sample ω from $\Lambda \setminus N_0$ (notice that the formulas which appear in the proof correspond to this ω). Since $\lim_{n\to\infty}(\gamma_{a(n,\hat{t})}/\gamma_n) = 1$, it can be deduced from Lemma 16 that there exist $n_0 > m_0 > \tau$ such that

$$\gamma_{m_0}^{\hat{p}} u(\theta_{m_0}) \leq 2\hat{C}_3, \tag{103}$$

$$\gamma_{n_0}^{\hat{p}} u(\theta_{n_0}) > \hat{C}_4, \tag{104}$$

$$\min_{m_0 < n \leq n_0} \gamma_n^{\hat{p}} u(\theta_n) > 2\hat{C}_3, \tag{105}$$

$$\max_{m_0 \leq n < n_0} \gamma_n^{\hat{p}} u(\theta_n) \leq \hat{C}_4, \tag{106}$$

and such that

$$(\gamma_{a(m_0,\hat{t})}/\gamma_{m_0})^{\hat{p}} \leq \min\{2, (1 - \hat{t}/\hat{C}_2)^{-1}\}. \tag{107}$$

Let $l_0 = a(m_0, \hat{t})$. As a direct consequence of Lemma 15 and (103), we get

$$\|\nabla f(\theta_{m_0})\|^2 \leq (4/\hat{t})\left(\varphi(u(\theta_{m_0})) + \gamma_{m_0}^{-\hat{p}}\right) \leq 12(\hat{C}_3/\hat{t})\gamma_{m_0}^{-\hat{p}}.$$

Consequently, Lemma 13 and (67) imply

$$\begin{aligned} u(\theta_n) - u(\theta_{m_0}) \leq |\phi_{m_0,n}| \leq & (\hat{t}/\hat{C}_1)(\gamma_{m_0}^{-2r} + \|\nabla f(\theta_{m_0})\|^2) \\ \leq & (\hat{t}/\hat{C}_1)\gamma_{m_0}^{-2r} + (12\hat{C}_3/\hat{C}_1)\gamma_{m_0}^{-\hat{p}} \leq \gamma_{m_0}^{-\hat{p}} \end{aligned}$$

for $m_0 \leq n \leq l_0$ (notice that $\hat{p} \leq 2r$, $\hat{t}/\hat{C}_1 \leq 1/2$, $\hat{C}_1 \geq 24\hat{C}_3$). Then, (103), (105) yield

$$u(\theta_{m_0}) \geq u(\theta_{m_0+1}) - \gamma_{m_0}^{-\hat{p}} \geq 2\hat{C}_3(\gamma_{m_0}/\gamma_{m_0+1})^{\hat{p}}\gamma_{m_0}^{-\hat{p}} - \gamma_{m_0}^{-\hat{p}} \geq (\hat{C}_3 - 1)\gamma_{m_0}^{-\hat{p}} \geq \gamma_{m_0}^{-\hat{p}}, \tag{108}$$

$$u(\theta_n) \leq u(\theta_{m_0}) + \gamma_{m_0}^{-\hat{p}} \leq (2\hat{C}_3 + 1)(\gamma_n/\gamma_{m_0})^{\hat{p}}\gamma_n^{-\hat{p}} \leq 6\hat{C}_3\gamma_n^{-\hat{p}} = \hat{C}_4\gamma_n^{-\hat{p}} \tag{109}$$

for $m_0 \leq n \leq l_0$ (notice that $(\gamma_n/\gamma_{m_0})^{\hat{p}} \leq (\gamma_{l_0}/\gamma_{m_0})^{\hat{p}} \leq 2$ for $m_0 \leq n \leq n_0$). Using (104), (109), we conclude $l_0 < n_0$.

In the rest of the proof, we consider separately the cases $\hat{\mu} < 2$ and $\hat{\mu} = 2$.

Case $\hat{\mu} < 2$: Owing to Lemma 14 (relations (87), (90)) and (103), (108), we have

$$\begin{aligned} v(\theta_{l_0}) \geq v(\theta_{m_0}) + \hat{t}/\hat{C}_2 \geq & (2\hat{C}_3)^{-1/\hat{p}}\,\gamma_{m_0} + (\gamma_{l_0} - \gamma_{m_0})/\hat{C}_2 \\ > & \min\{(2\hat{C}_3)^{-1/\hat{p}}, \hat{C}_2^{-1}\}\gamma_{l_0} \\ = & (2\hat{C}_3)^{-1/\hat{p}}\,\gamma_{l_0} \end{aligned}$$

(notice that $(2\hat{C}_3)^{1/\hat{p}} > \hat{C}_2$). Therefore,

$$u(\theta_{l_0}) = (v(\theta_{l_0}))^{-\hat{p}} < 2\hat{C}_3\gamma_{l_0}^{-\hat{p}}.$$

However, this directly contradicts (105) and the fact that $m_0 < l_0 < n_0$. Thus, (102) holds when $\hat{\mu} < 2$.

Case $\hat{\mu} = 2$*:* Using Lemma 14 (relations (86),(89)) and (108), we get

$$u(\theta_{l_0}) \leq (1 - \hat{t}/\hat{C}_2)u(\theta_{m_0}) \leq 2\hat{C}_3(1 - \hat{t}/\hat{C}_2)(\gamma_{l_0}/\gamma_{m_0})^{\hat{p}}\gamma_{l_0}^{-\hat{p}} \leq 2\hat{C}_3\gamma_{l_0}^{-\hat{p}}.$$

However, this is impossible due to (105) and the fact that $m_0 < l_0 < n_0$. Hence, (102) holds in the case $\hat{\mu} = 2$, too. ■

Lemma 18: Suppose that Assumptions 1 – 4 hold. Then,

$$\|\theta_{a(n,\hat{t})} - \theta_n\| \leq 2\gamma_n^{\hat{q}+1}(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + 6\gamma_n^{-(\hat{q}+1)} \tag{110}$$

on $\Lambda \setminus N_0$ for $n > \tau$.

Proof: Let ω be an arbitrary sample from $\Lambda \setminus N_0$, while $n > \max\{\sigma, \tau\}$ is an arbitrary integer (notice that all formulas which appear in the proof correspond to these ω, n). To show (110), we consider separately the cases $\|\nabla f(\theta_n)\| \geq \gamma_n^{-(\hat{q}+1)}$ and $\|\nabla f(\theta_n)\| < \gamma_n^{-(\hat{q}+1)}$.

Case $\|\nabla f(\theta_n)\| \geq \gamma_n^{-(\hat{q}+1)}$*:* Due to Lemma 13, we have

$$\|\nabla f(\theta_n)\|\|\theta_{a(n,\hat{t})} - \theta_n\| \leq 2(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + 2(\hat{t}/\hat{C}_1)\gamma_n^{-2r}. \tag{111}$$

On the other side, since $\|\nabla f(\theta_n)\| \geq \gamma_n^{-(\hat{q}+1)} \geq \gamma_n^{-r}$ (notice that $\hat{q} + 1 = \min\{(\hat{p}+1)/2, r\} \leq r$), Lemma 14 (relations (85), (88)) implies

$$u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/4)\|\nabla f(\theta_n)\|^2 < 0,$$

i.e., $u(\theta_n) - u(\theta_{a(n,\hat{t})}) > 0$. Then, (111) yields

$$\begin{aligned}\|\theta_{a(n,\hat{t})} - \theta_n\| \leq& 2(u(\theta_n) - u(\theta_{a(n,\hat{t})}))\|\nabla f(\theta_n)\|^{-1} + 2(\hat{t}/\hat{C}_1)\gamma_n^{-2r}\|\nabla f(\theta_n)\|^{-1}\\ \leq& 2\gamma_n^{\hat{q}+1}(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + \gamma_n^{-2r+(\hat{q}+1)}\\ \leq& 2\gamma_n^{\hat{q}+1}(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + \gamma_n^{-(\hat{q}+1)}\end{aligned}$$

(notice that $\hat{t}/\hat{C}_1 \leq 1/2$; also notice that $\hat{q} + 1 \leq r$, which implies $2r - (\hat{q}+1) \geq \hat{q} + 1$). Hence, (110) is true when $\|\nabla f(\theta_n)\| \geq \gamma_n^{-(\hat{q}+1)}$.

Case $\|\nabla f(\theta_n)\| < \gamma_n^{-(\hat{q}+1)}$*:* Using Lemma 13 and (67), we get

$$\begin{aligned}|u(\theta_{a(n,\hat{t})}) - u(\theta_n)| \leq& (\gamma_{a(n,\hat{t})} - \gamma_n)\|\nabla f(\theta_n)\|^2 + |\phi_{n,a(n,\hat{t})}|\\ \leq& \hat{t}\|\nabla f(\theta_n)\|^2 + (\hat{t}/\hat{C}_1)(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2)\\ \leq& 2\gamma_n^{-2(\hat{q}+1)}\end{aligned}$$

(notice that $\hat{q}+1 \leq r < 2r$ and $\hat{t}/\hat{C}_1 \leq 1/2$). On the other side, owing to Lemma 13 and (66), we have

$$\begin{aligned}\|\theta_{a(n,\hat{t})} - \theta_n\| \leq &(\gamma_{a(n,\hat{t})} - \gamma_n)\|\nabla f(\theta_n)\| + \|\varepsilon_{n,a(n,\hat{t})}\| \\ \leq &\hat{t}\|\nabla f(\theta_n)\| + (\hat{t}/\hat{C}_1)(\gamma_n^{-r} + \|\nabla f(\theta_n)\|) \\ \leq &2\gamma_n^{-(\hat{q}+1)}\end{aligned}$$

(notice that $\hat{q}+1 \leq r$). Consequently,

$$\begin{aligned}\|\theta_{a(n,\hat{t})} - \theta_n\| \leq &2\gamma_n^{\hat{q}+1}(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + 2\gamma_n^{\hat{q}+1}|u(\theta_n) - u(\theta_{a(n,\hat{t})})| + 2\gamma_n^{-(\hat{q}+1)} \\ \leq &2\gamma_n^{\hat{q}+1}(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + 6\gamma_n^{-(\hat{q}+1)}.\end{aligned}$$

Thus, (110) holds in the case $\|\nabla f(\theta_n)\| < \gamma_n^{-(\hat{q}+1)}$. ■

Lemma 19: Suppose that Assumptions 1 – 4 hold. Then, there exists a random quantity $\hat{C}_5$ such that $1 \leq \hat{C}_5 < \infty$ everywhere and such that

$$\limsup_{n\to\infty} \gamma_n^{\hat{q}} \max_{k\geq n} \|\theta_k - \theta_n\| \leq \hat{C}_5 \tag{112}$$

on $\Lambda \setminus N_0$.

Proof: Let $\tilde{C} = 9\hat{C}_4(\hat{q}+1)$ and $\hat{C}_5 = 20\tilde{C}\hat{t}^{-1}(1+1/\hat{q})$, while ω is an arbitrary sample from $\Lambda \setminus N_0$ (notice that all formulas which follow in the proof correspond to this ω).

As a consequence of Lemmas 15 and 17, we get

$$\limsup_{n\to\infty} \gamma_n^{\hat{p}}|u(\theta_n)| \leq \hat{C}_4, \tag{113}$$

$$\limsup_{n\to\infty} \gamma_n^{\hat{p}}\|\nabla f(\theta_n)\|^2 \leq 8\hat{C}_4/\hat{t}. \tag{114}$$

Since $\gamma_{a(n,\hat{t})} - \gamma_n = \hat{t} + O(\alpha_{a(n,\hat{t})})$ for $n \to \infty$, and

$$(1 - \hat{t}/\gamma_n)^{\hat{q}+1} = 1 - \hat{t}(\hat{q}+1)\gamma_n^{-1} + o(\gamma_n^{-1})$$

for $n \to \infty$, we conclude from (113), (114) that there exists $n_0 > \max\{\sigma, \tau\}$ (depending on ω) such that $|u(\theta_n)| \leq 2\hat{C}_4\gamma_n^{-\hat{p}}$, $\|\nabla f(\theta_n)\| \leq (4\hat{C}_4/\hat{t})\gamma_n^{-\hat{p}/2}$, $\gamma_{a(n,\hat{t})} - \gamma_n \geq \hat{t}/2$ and

$$(1 - \hat{t}/\gamma_n)^{\hat{q}+1} \geq 1 - (\hat{q}+1)\gamma_n^{-1} \tag{115}$$

for $n \geq n_0$. Then, (66) and Lemma 13 imply

$$\begin{aligned}\|\theta_k - \theta_n\| \leq &(\gamma_k - \gamma_n)\|\nabla f(\theta_n)\| + \|\varepsilon_{n,k}\| \\ \leq &\hat{t}\|\nabla f(\theta_n)\| + (\hat{t}/\hat{C}_1)(\gamma_n^{-r} + \|\nabla f(\theta_n)\|) \\ \leq &8\hat{C}_4\gamma_n^{-\hat{p}/2} + \gamma_n^{-r} \\ \leq &\tilde{C}\gamma_n^{-\hat{q}}\end{aligned} \tag{116}$$

for $n_0 \leq n \leq k \leq a(n,\hat{t})$ (notice that $\hat{q} < \min\{\hat{p}/2, r\}$).

Let $\{n_k\}_{k\geq 0}$ be recursively defined as $n_{k+1}=a(n_k,\hat{t})$ for $k\geq 0$. Due to Lemma 18, we have

$$\begin{aligned}\|\theta_{n_l}-\theta_{n_k}\|\leq\sum_{i=k}^{l-1}\|\theta_{n_{i+1}}-\theta_{n_i}\|\leq & 6\sum_{i=k}^{l-1}\gamma_{n_i}^{-(\hat{q}+1)}+2\sum_{i=k}^{l-1}\gamma_{n_i}^{\hat{q}+1}(u(\theta_{n_i})-u(\theta_{n_{i+1}})) \\ \leq & 6\sum_{i=k}^{l-1}\gamma_{n_i}^{-(\hat{q}+1)}+2\sum_{i=k+1}^{l}(\gamma_{n_i}^{\hat{q}+1}-\gamma_{n_{i-1}}^{\hat{q}+1})|u(\theta_{n_i})| \\ & +2\gamma_{n_l}^{\hat{q}+1}|u(\theta_{n_l})|+2\gamma_{n_k}^{\hat{q}+1}|u(\theta_{n_k})| \end{aligned}\tag{117}$$

for $l\geq k\geq 0$. As

$$\gamma_{n_i}^{\hat{q}+1}-\gamma_{n_{i-1}}^{\hat{q}+1}=\gamma_{n_i}^{\hat{q}+1}\left(1-\left(1-(\gamma_{n_i}-\gamma_{n_{i-1}})/\gamma_{n_i}\right)^{\hat{q}+1}\right)\leq\gamma_{n_i}^{\hat{q}+1}\left(1-\left(1-\hat{t}/\gamma_{n_i}\right)^{\hat{q}+1}\right)\leq(\hat{q}+1)\gamma_{n_i}^{\hat{q}}$$

for $i\geq 0$ (use (115)), we get

$$\sum_{i=k+1}^{l}(\gamma_{n_i}^{\hat{q}+1}-\gamma_{n_{i-1}}^{\hat{q}+1})|u(\theta_{n_i})|\leq 2\hat{C}_4(\hat{q}+1)\sum_{i=k}^{\infty}\gamma_{n_i}^{-\hat{p}+\hat{q}}\leq\tilde{C}\sum_{i=k}^{\infty}\gamma_{n_i}^{-(\hat{q}+1)}\tag{118}$$

for $l>k\geq 0$ (notice that $\hat{p}-\hat{q}\geq(\hat{p}+1)/2\geq\hat{q}+1$). Since

$$\gamma_{n_l}=\gamma_{n_k}+\sum_{i=k}^{l-1}(\gamma_{n_{i+1}}-\gamma_{n_i})\geq\gamma_{n_k}+(\hat{t}/2)(l-k)$$

for $l>k\geq 0$ (notice that $\gamma_{a(n,\hat{t})}-\gamma_n\geq\hat{t}/2$) for $n\geq n_0$), we have

$$\begin{aligned}\sum_{i=k}^{\infty}\gamma_{n_i}^{-(\hat{q}+1)}\leq & \sum_{i=0}^{\infty}(\gamma_{n_k}+\hat{t}i/2)^{-(\hat{q}+1)} \\ \leq & \gamma_{n_k}^{-(\hat{q}+1)}+\int_0^{\infty}(\gamma_{n_k}+\hat{t}u/2)^{-(\hat{q}+1)}du \\ = & \gamma_{n_k}^{-(\hat{q}+1)}+2\hat{t}^{-1}\hat{q}^{-1}\gamma_{n_k}^{-\hat{q}} \\ \leq & (1+2\hat{t}^{-1}\hat{q}^{-1})\gamma_{n_k}^{-\hat{q}}\end{aligned}$$

for $k\geq 0$. Consequently, (117) and (118) imply

$$\|\theta_{n_l}-\theta_{n_k}\|\leq(6+2\tilde{C})\sum_{i=k}^{\infty}\gamma_{n_i}^{-(\hat{q}+1)}+4\hat{C}_4\gamma_{n_k}^{-\hat{p}+\hat{q}+1}+4\hat{C}_4\gamma_{n_l}^{-\hat{p}+\hat{q}+1}\leq 16\tilde{C}(1+\hat{t}^{-1}\hat{q}^{-1})\gamma_{n_k}^{-\hat{q}}\tag{119}$$

for $l\geq k\geq 0$ (notice that $\hat{p}-(\hat{q}+1)\geq(\hat{p}-1)/2\geq\hat{q}$). Using (116) and (119), we get

$$\begin{aligned}\|\theta_k-\theta_n\|\leq & \|\theta_k-\theta_{n_j}\|+\|\theta_{n_j}-\theta_{n_i}\|+\|\theta_{n_i}-\theta_n\| \\ \leq & \tilde{C}\gamma_k^{-\hat{q}}+\tilde{C}\gamma_n^{-\hat{q}}+16\tilde{C}(1+\hat{t}^{-1}\hat{q}^{-1})\gamma_{n_i}^{-\hat{q}} \\ \leq & \hat{C}_5\gamma_n^{-\hat{q}}\end{aligned}$$

for $k\geq n\geq n_0$, $j\geq i\geq 1$ satisfying $n_{i-1}\leq n<n_i$, $n_{j-1}\leq k<n_j$. Then, it is obvious that (112) is true. ■

Proof of Theorems 2 and 3: Owing to Lemmas 11 and 19, we have that on $\Lambda\setminus N_0$, $\hat{\theta}=\lim_{n\to\infty}\theta_n$ exists and satisfies $\nabla f(\hat{\theta})=0$. Consequently, $\hat{Q}\subseteq\{\theta\in\mathbb{R}^{d_\theta}:\|\theta-\hat{\theta}\|\leq\delta_{\hat{\theta}}\}$ on $\Lambda\setminus N_0$. Thus, random quantities $\hat{p}$, $\hat{q}$ defined in this subsection coincide with $\hat{p}$, $\hat{q}$ introduced in Theorem 3 (see the remark after Corollary 5). Then, Lemmas 15, 17, 19 imply that (4) is true on $\Lambda\setminus N_0$. ■

V. Proof of Propositions 1 – 4

Proof of Proposition 1: Owing to Conditions (i), (ii) of the proposition, for any compact set $Q \subset \Theta$, there exists a real number $\varepsilon_Q \in (0,1)$ such that

$$\varepsilon_Q \leq r_\theta(y|x',x) \leq \varepsilon_Q^{-1} \tag{120}$$

for all $\theta \in Q$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Hence, Assumption 3 is satisfied. On the other side, Condition (ii) implies that $r_\theta(y|x',x)$ has an (complex-valued) analytical continuation $\hat{r}_\eta(y|x',x)$ with the following properties:

a) $\hat{r}_\eta(y|x',x)$ maps $(\eta, x, x', y) \in \mathbb{C}^{d_\theta} \times \mathcal{X} \times \mathcal{X} \times \mathcal{Y}$ into $\mathbb{C}$.
b) $\hat{r}_\theta(y|x',x) = r_\theta(y|x',x)$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.
c) For any compact set $Q \subset \Theta$, there exists a real number $\tilde{\delta}_Q \in (0,1)$ such that $\hat{r}_\eta(y|x',x)$ is analytical in η on $V_{\tilde{\delta}_Q}(Q)$ for each $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.

Relying on $\hat{r}_\eta(y|x',x)$, we define quantities $\hat{R}_\eta(y)$, $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$. More specifically, for $\eta \in \mathbb{C}^{d_\theta}$, $y \in \mathcal{Y}$, $\hat{R}_\eta(y)$ is an $N_x \times N_x$ matrix whose (i,j) entry is $\hat{r}_\eta(y|i,j)$, while

$$\hat{\phi}_\eta(w,y) = \begin{cases} \log(e^T \hat{R}_\eta(y) w), & \text{if } e^T \hat{R}_\eta(y) w \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{121}$$

$$\hat{G}_\eta(w,y) = \begin{cases} \hat{R}_\eta(y) w / (e^T \hat{R}_\eta(y) w), & \text{if } e^T \hat{R}_\eta(y) w \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{122}$$

for $\eta \in \mathbb{C}^{d_\theta}$, $y \in \mathcal{Y}$, $w \in \mathbb{C}^{N_x}$.

Let $Q \subset \Theta$ be an arbitrary compact set. Since $e^T R_\theta(y) u \geq N_x \varepsilon_Q$ for all $\theta \in Q$, $y \in \mathcal{Y}$, $u \in \mathcal{P}^{N_x}$ (due to (120)), we conclude that there exists a real number $\delta_Q \in (0, \tilde{\delta}_Q)$ such that $|e^T \hat{R}_\eta(y) w| \geq N_x \varepsilon_Q / 2$ for all $\eta \in V_{\delta_Q}(Q)$, $w \in V_{\delta_Q}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$. Therefore, $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are analytical in (η, w) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x})$ for any $y \in \mathcal{Y}$. Consequently, $|\hat{\phi}_\eta(w,y)|$, $\|\hat{G}_\eta(w,y)\|$ are uniformly bounded in (η, w, y) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x}) \times \mathcal{Y}$. Thus, Assumption 4 is satisfied, too. ■

Proof of Proposition 2: Conditions (i), (ii) of the proposition imply that for any compact set $Q \subset \Theta$, there exists a real number $\varepsilon_Q \in (0,1)$ such that $\varepsilon_Q \leq r_\theta(y|x',x) \leq \varepsilon_Q^{-1}$ for all $\theta \in Q$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Thus, Assumption 3 holds. On the other side, as a result of Condition (ii), $r_\theta(y|x',x)$ has an (complex-valued) analytical continuation $\hat{r}_\eta(z|x',x)$ with the following properties:

a) $\hat{r}_\eta(z|x',x)$ maps $(\eta, x, x', z) \in \mathbb{C}^{d_\theta} \times \mathcal{X} \times \mathcal{X} \times \mathbb{C}^{d_y}$ into $\mathbb{C}$.
b) $\hat{r}_\theta(y|x',x) = r_\theta(y|x',x)$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$.
c) For any compact set $Q \subset \Theta$, there exists a real number $\tilde{\delta}_Q \in (0,1)$ such that $\hat{r}_\eta(z|x',x)$ is analytical in (η, z) on $V_{\tilde{\delta}_Q}(Q) \times V_{\tilde{\delta}_Q}(\mathcal{Y})$ for each $x, x' \in \mathcal{X}$.

Relying on $\hat{r}_\eta(y|x',x)$, we define quantities $\hat{R}_\eta(y)$, $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ in the same way as in the proof of Proposition 1. More specifically, for $\eta \in \mathbb{C}^{d_\theta}$, $y \in \mathcal{Y}$, $\hat{R}_\eta(y)$ is an $N_x \times N_x$ matrix whose (i,j) entry is $\hat{r}_\eta(y|i,j)$, while $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are defined by (121), (122) for $\eta \in \mathbb{C}^{d_\theta}$, $y \in \mathcal{Y}$, $w \in \mathbb{C}^{N_x}$.

Let $Q \subset \Theta$ be an arbitrary compact set. As $N_x \varepsilon_Q \leq e^T R_\theta(y) u \leq N_x \varepsilon_Q^{-1}$ for any $\theta \in Q$, $y \in \mathcal{Y}$, $u \in \mathcal{P}^{N_x}$, we have that there exists a real number $\delta_Q \in (0, \tilde{\delta}_Q)$ such that $N_x \varepsilon_Q / 2 \leq |e^T \hat{R}_\eta(y) w| \leq 2 N_x \varepsilon_Q^{-1}$ for all $\eta \in V_{\delta_Q}(Q)$, $w \in V_{\delta_Q}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$ (notice that $|e^T \hat{R}_\eta(y) w|$ is analytical in (η, w, y) on $V_{\tilde{\delta}_Q}(Q) \times V_{\tilde{\delta}_Q}(\mathcal{P}^{N_x}) \times \mathcal{Y}$). Therefore, $\hat{\phi}_\eta(w, y)$, $\hat{G}_\eta(w, y)$ are analytical in (η, w) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x})$ for any $y \in \mathcal{Y}$. Moreover, $|\hat{\phi}_\eta(w, y)|$, $\|\hat{G}_\eta(w, y)\|$ are uniformly bounded in (η, w, y) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x}) \times \mathcal{Y}$. Hence, Assumption 4 holds, too. ■

Proof of Proposition 3: For $\alpha \in \mathcal{A}$, $\beta = [\beta_1 \cdots \beta_{N_\beta}]^T \in \mathcal{B}$, $x, x' \in \mathcal{X}$, let $g_\theta^k(x'|x) = \beta_{x',k} p_\alpha(x'|x)$. Then, we have

$$r_\theta(y|x', x) = \sum_{k=1}^{N_\beta} f_k(y|x') g_\theta^k(x'|x)$$

for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. We also have that for any compact set $Q \subset \Theta$, there exists a real number $\varepsilon_Q \in (0, 1)$ such that $\varepsilon_Q \leq g_\theta^k(x'|x) \leq \varepsilon_Q^{-1}$ for each $\theta \in Q$, $x, x' \in \mathcal{X}$, $1 \leq k \leq N_\beta$. Consequently,

$$\varepsilon_Q \sum_{k=1}^{N_\beta} f_k(y|x') \leq r_\theta(y|x', x) \leq \varepsilon_Q^{-1} \sum_{k=1}^{N_\beta} f_k(y|x')$$

for all $\theta \in Q$, $x, x' \in \mathcal{X}$ and any compact set $Q \subset \Theta$. Hence, Assumption 3 holds (set $s_\theta(y|x) = \sum_{k=1}^{N_\beta} f_k(y|x)$). On the other side, Condition (i) implies that for each $1 \leq k \leq N_\beta$, $g_\theta^k(x'|x)$ has an (complex-valued) analytical continuation $\hat{g}_\eta^k(x'|x)$ with the following properties:

a) $\hat{g}_\eta(x'|x)$ maps $(\eta, x, x') \in \mathbb{C}^{d_\theta} \times \mathcal{X} \times \mathcal{X}$ into $\mathbb{C}$.
b) $\hat{g}_\theta^k(x'|x) = g_\theta^k(x'|x)$ for all $\theta \in \Theta$, $x, x' \in \mathcal{X}$.
c) For any compact set $Q \subset \Theta$, there exists a real number $\tilde{\delta}_Q \in (0, 1)$ such that $\hat{g}_\eta^k(x'|x)$ is analytical in η on $V_{\tilde{\delta}_Q}(Q)$ for each $x, x' \in \mathcal{X}$.

Relying on $\hat{g}_\eta^k(x'|x)$, we define some new quantities. More specifically, for $\eta \in \mathbb{C}^{d_\theta}$, $w = [w_1 \cdots w_{N_x}]^T \in \mathbb{C}^{N_x}$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, let

$$\hat{r}_\eta(y|x', x) = \sum_{k=1}^{N_\beta} f_k(y|x') \hat{g}_\eta^k(x'|x),$$

$$\hat{h}_{\eta,w}^k(x') = \sum_{x'' \in \mathcal{X}} \hat{g}_\eta^k(x'|x'') w_{x''},$$

while $\hat{R}_\eta(y)$ is an $N_x \times N_x$ matrix whose (i, j) entry is $\hat{r}_\eta(y|i, j)$. Moreover, let $\hat{\phi}_\eta(w, y)$, $\hat{G}_\eta(w, y)$ be defined for $\eta \in \mathbb{C}^{d_\theta}$, $w \in \mathbb{C}^{N_x}$, $y \in \mathcal{Y}$ in the same way as in (121), (122).

Let $Q \subset \Theta$ be arbitrary compact set. Since

$$\varepsilon_Q \leq \sum_{x \in \mathcal{X}} g_\theta^k(x'|x) u_x \leq \varepsilon_Q^{-1}$$

for all $\theta \in Q$, $u = [u_1 \cdots u_{N_x}]^T \in \mathcal{P}^{N_x}$, $x, x' \in \mathcal{X}$, $1 \leq k \leq N_\beta$, we deduce that there exists a real number $\delta_Q \in (0, \tilde{\delta}_Q)$ such that $\text{Re}\{\hat{h}_{\eta,w}^k(x')\} \geq \varepsilon_Q / 2$, $|\hat{h}_{\eta,w}^k(x')| \leq 2\varepsilon_Q^{-1}$ for all $\eta \in V_{\delta_Q}(Q)$, $w \in V_{\delta_Q}(\mathcal{P}^{N_x})$, $x' \in \mathcal{X}$,

$1 \leq k \leq N_\beta$. Consequently,

$$|e^T \hat{R}_\eta(y)w| \geq |\mathrm{Re}\{e^T \hat{R}_\eta(y)w\}| = \sum_{x' \in \mathcal{X}} \sum_{k=1}^{N_\beta} f_k(y|x')\mathrm{Re}\{\hat{h}^k_{\eta,w}(x')\} \geq (\varepsilon_Q/2)\psi(y) > 0,$$

$$\max\{\|\hat{R}_\eta(y)w\|, |e^T \hat{R}_\eta(y)w|\} \leq \sum_{x' \in \mathcal{X}} \sum_{k=1}^{N_\beta} f_k(y|x')|\hat{h}^k_{\eta,w}(x')| \leq 2\varepsilon_Q^{-1}\psi(y)$$

for all $\eta \in V_{\delta_Q}(Q)$, $w \in V_{\delta_Q}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$. Therefore, $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are analytical in (η, w) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x})$ for each $y \in \mathcal{Y}$. Moreover,

$$\|\hat{G}_\eta(w,y)\| \leq 4\varepsilon_Q^{-2},$$

$$|\hat{\phi}_\eta(w,y)| \leq |\log|e^T \hat{R}_\eta(y)w|| + 2\pi \leq |\log\psi(y)| + \log(2\varepsilon_Q^{-1}) + 2\pi$$

for all $\eta \in V_{\delta_Q}(Q)$, $w \in V_{\delta_Q}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$. Then, it is clear that Assumption 4 holds, too. ■

Lemma 20: Let the conditions of Proposition 4 hold. Then, $\phi_\theta(u,y)$, $G_\theta(u,y)$ have (complex-valued) analytical continuations $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ (respectively) with the following properties:

i) $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ map $(\eta, w, y) \in \mathbb{C}^{d_\theta} \times \mathbb{C}^{N_x} \times \mathcal{Y}$ into $\mathbb{C}$, $\mathbb{C}^{N_x}$ (respectively).

ii) $\hat{\phi}_\theta(u,y) = \phi_\theta(u,y)$, $\hat{G}_\theta(u,y) = G_\theta(u,y)$ for all $\theta \in \Theta$, $u \in \mathcal{P}^{N_x}$, $y \in \mathcal{Y}$.

iii) For each $\theta \in \Theta$, there exist real numbers $\delta_\theta \in (0,1)$, $K_\theta \in [1, \infty)$ such that $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are analytical in (η, w) on $V_{\delta_\theta}(\theta) \times V_{\delta_\theta}(\mathcal{P}^{N_x})$ for any $y \in \mathcal{Y}$, and such that

$$|\hat{\phi}_\eta(w,y)| \leq K_\theta(1 + y^2),$$

$$\|\hat{G}_\eta(w,y)\| \leq K_\theta$$

for all $\eta \in V_{\delta_\theta}(\theta)$, $w \in V_{\delta_\theta}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$.

Proof: Due to Condition (i) of Proposition 4, $p_\alpha(x'|x)$ has an (complex-valued) analytical continuation $\hat{p}_a(x'|x)$ with the following properties

a) $\hat{p}_a(x'|x)$ maps $(a, x, x') \in \mathbb{C}^{d_\alpha} \times \mathcal{X} \times \mathcal{X}$ into $\mathbb{C}$.

b) $\hat{p}_\alpha(x'|x) = p_\alpha(x'|x)$ for all $\alpha \in \mathcal{A}$, $x, x' \in \mathcal{X}$.

c) For any $\alpha \in \mathcal{A}$, there exists a real number $\tilde{\delta}_\alpha \in (0,1)$ such that $\hat{p}_a(x'|x)$ is analytical in a on $V_{\tilde{\delta}_\alpha}(\alpha)$ for each $x, x' \in \mathcal{X}$.

On the other side, the analytical continuation $\hat{q}_b(y|x)$ of $q_\beta(y|x)$ is defined by

$$\hat{q}_b(y|x) = \sqrt{l_x/\pi}\exp(-l_x(y - m_x)^2),$$

for $b = [l_1 \cdots l_{N_x}\ m_1 \cdots m_{N_x}]^T \in \mathbb{C}^{2N_x}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

Let $\hat{r}_\eta(y|x', x) = \hat{q}_b(y|x')\hat{p}_a(x'|x)$ for $a \in \mathbb{C}^{d_\alpha}$, $b \in \mathbb{C}^{2N_x}$, $\eta = [a^T\ b^T]^T$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$. Moreover, for $\eta \in \mathbb{C}^{d_\theta}$, $y \in \mathcal{Y}$, $\hat{R}_\eta(y)$ is an $N_x \times N_x$ matrix whose (i,j) entry is $\hat{r}_\eta(y|i,j)$, while $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are defined for $\eta \in \mathbb{C}^{d_\theta}$, $w \in \mathbb{C}^{N_x}$, $y \in \mathcal{Y}$ in the same way as in (121), (122).

Let α, $\beta = [\lambda_1 \cdots \lambda_{N_x} \, \mu_1 \, \mu_{N_x}]^T$ be arbitrarily vectors from $\mathcal{A}$, $\mathcal{B}$ (respectively), while $\theta = [\alpha^T \, \beta^T]^T$. Obviously, it can be assumed without loss of generality that $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_{N_x}$. Since

$$\sum_{x \in \mathcal{X}} p_\alpha(x'|x) u_x > 0$$

for all $x' \in \mathcal{X}$, $u = [u_1 \cdots u_{N_x}]^T \in \mathcal{P}^{N_x}$, there exist real numbers $\tilde{\delta}_{1,\theta}, \tilde{\varepsilon}_\theta \in (0,1)$ such that $\hat{R}_\eta(y)$ is analytical in η on $V_{\tilde{\delta}_{1,\theta}}(\theta)$ for any $y \in \mathcal{Y}$, and such that

$$\text{Re}\left\{ \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right\} \geq \tilde{\varepsilon}_\theta, \tag{123}$$

$$\left| \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| \leq \tilde{\varepsilon}_\theta^{-1}, \tag{124}$$

$$\min\{\text{Re}\{l_1\}, \text{Re}\{l_{x'} - l_1\}\} \geq \tilde{\varepsilon}_\theta,$$

$$\max\{|l_{x''}|, |m_{x''}|\} \leq \tilde{\varepsilon}_\theta^{-1}$$

for all $a \in V_{\tilde{\delta}_{1,\theta}}(\alpha)$, $b = [l_1 \cdots l_{N_x} \; m_1 \cdots m_{N_x}]^T \in V_{\tilde{\delta}_{1,\theta}}(\beta)$, $w = [w_1 \cdots w_{N_x}]^T \in V_{\tilde{\delta}_{1,\theta}}(\mathcal{P}^{N_x})$, $x' \in \mathcal{X} \setminus \{1\}$, $x'' \in \mathcal{X}$. Therefore, we have

$$\begin{aligned} |\hat{q}_b(y|x)| =& \sqrt{|l_x|/\pi} \, |\exp(-\text{Re}\{l_x\} y^2 + 2\text{Re}\{l_x m_x\} y - \text{Re}\{l_x m_x^2\})| \\ \leq& \sqrt{|l_x|/\pi} \, \exp(-\text{Re}\{l_x\} y^2 + 2|l_x||m_x||y| + |l_x||m_x|^2) \\ \leq& (1/\sqrt{\pi \varepsilon_\theta}) \, \exp(-\tilde{\varepsilon}_\theta y^2 + 2\tilde{\varepsilon}_\theta^{-2}|y| + \tilde{\varepsilon}_\theta^{-3}) \end{aligned}$$

for any $b = [l_1 \cdots l_{N_x} \; m_1 \cdots m_{N_x}]^T \in V_{\tilde{\delta}_{1,\theta}}(\beta)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. We also have

$$\begin{aligned} \left| \frac{\hat{q}_b(y|x)}{\hat{q}_b(y|1)} \right| =& \sqrt{|l_x|/|l_1|} \, |\exp(-\text{Re}\{l_x - l_1\} y^2 + 2\text{Re}\{l_x m_x - l_1 m_1\} y - \text{Re}\{l_x m_x^2 - l_1 m_1^2\})| \\ \leq& \sqrt{|l_x|/|l_1|} \, \exp(-\text{Re}\{l_x - l_1\} y^2 + 2(|l_x||m_x| + |l_1||m_1|)|y| + |l_x||m_x|^2 + |l_1||m_1|^2) \\ \leq& \varepsilon_\theta^{-1} \exp(-\tilde{\varepsilon}_\theta y^2 + 4\tilde{\varepsilon}_\theta^{-2}|y| + 2\tilde{\varepsilon}_\theta^{-3}) \end{aligned}$$

for all $b = [l_1 \cdots l_{N_x} \; m_1 \cdots m_{N_x}]^T \in V_{\tilde{\delta}_{1,\theta}}(\beta)$, $x \in \mathcal{X} \setminus \{1\}$, $y \in \mathcal{Y}$. Consequently, there exists a real number $\tilde{C}_\theta \in [1, \infty)$ such that

$$\left| \frac{\hat{q}_b(y|x)}{\hat{q}_b(y|1)} \right| \leq \tilde{C}_\theta, \tag{125}$$

$$|\log |\hat{q}_b(y|x)|| \leq \tilde{C}_\theta (1 + y^2) \tag{126}$$

for all $b \in V_{\tilde{\delta}_{1,\theta}}(\beta)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and such that

$$\left| \frac{\hat{q}_b(y|x)}{\hat{q}_b(y|1)} \right| \leq 2^{-1} N_x^{-1} \tilde{\varepsilon}_\theta^2 \tag{127}$$

for any $b \in V_{\tilde{\delta}_{1,\theta}}(\beta)$, $x \in \mathcal{X} \setminus \{1\}$, $y \in [-\tilde{C}_\theta, \tilde{C}_\theta]^c$ (to show that (127) holds for all sufficiently large $|y|$, notice that $\lim_{|y| \to \infty} \hat{q}_b(y|x)/\hat{q}_b(y|1) = 0$ for $x \neq 1$). As $\hat{q}_b(y|x)/q_\beta(y|x)$ is uniformly continuous in (b, y) on

$V_{\tilde{\delta}_{1,\theta}}(\beta) \times [-\tilde{C}_\theta, \tilde{C}_\theta]$ and $\lim_{b \to \beta} \hat{q}_b(y|x)/q_\beta(y|x) = 1$ for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, there also exists a real number $\tilde{\delta}_{2,\theta} \in (0,1)$ such that

$$\left| \frac{\hat{q}_b(y|x)}{q_\beta(y|x)} - 1 \right| \leq 2^{-1} \tilde{\varepsilon}_\theta^2, \tag{128}$$

$$\left| \frac{\hat{q}_b(y|x)}{q_\beta(y|x)} \right| \leq 2 \tag{129}$$

for all $b \in V_{\tilde{\delta}_{2,\theta}}(\beta)$, $x \in \mathcal{X}$, $y \in [-\tilde{C}_\theta, \tilde{C}_\theta]$.

Let $\delta_\theta = \min\{\tilde{\delta}_{1,\theta}, \tilde{\delta}_{2,\theta}\}$, $K_\theta = 8 N_x \tilde{C}_\theta \tilde{\varepsilon}_\theta^{-2}$. As a result of (124), (125), we have

$$\max\{\|\hat{R}_\eta(y)w\|, |e^T \hat{R}_\eta(y)w|\} \leq \sum_{x' \in \mathcal{X}} |\hat{q}_b(y|x')| \left| \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| \leq N_x \tilde{C}_\theta \tilde{\varepsilon}_\theta^{-1} |\hat{q}_b(y|1)| \tag{130}$$

for all $a \in V_{\delta_\theta}(\alpha)$, $b \in V_{\delta_\theta}(\beta)$, $\eta = [a^T b^T]^T$, $y \in \mathcal{Y}$, $w = [w_1 \cdots w_{N_x}]^T \in V_{\delta_\theta}(\mathcal{P}^{N_x})$. Using (123), (124), (127), we get

$$\begin{aligned} |e^T \hat{R}_\eta(y)w| =& |\hat{q}_b(y|1)| \left| \sum_{x' \in \mathcal{X}} \frac{\hat{q}_b(y|x')}{\hat{q}_b(y|1)} \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| \\ \geq& |\hat{q}_b(y|1)| \left(\mathrm{Re}\left\{ \sum_{x \in \mathcal{X}} \hat{p}_a(1|x) w_x \right\} - \sum_{x' \in \mathcal{X} \setminus \{1\}} \left| \frac{\hat{q}_b(y|x')}{\hat{q}_b(y|1)} \right| \left| \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| \right) \\ \geq& 2^{-1} \tilde{\varepsilon}_\theta |\hat{q}_b(y|1)| \end{aligned} \tag{131}$$

for all $a \in V_{\delta_\theta}(\alpha)$, $b \in V_{\delta_\theta}(\beta)$, $\eta = [a^T b^T]^T$, $y \in [-\tilde{C}_\theta, \tilde{C}_\theta]^c$, $w = [w_1 \cdots w_{N_x}]^T \in V_{\delta_\theta}(\mathcal{P}^{N_x})$. Combining (123), (124), (128), (129), we obtain

$$\begin{aligned} |e^T \hat{R}_\eta(y)w| \geq& \left| \sum_{x' \in \mathcal{X}} q_\beta(y|x') \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| - \left| \sum_{x' \in \mathcal{X}} (\hat{q}_b(y|x') - q_\beta(y|x')) \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| \\ \geq& \sum_{x' \in \mathcal{X}} q_\beta(y|x') \mathrm{Re}\left\{ \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right\} - \sum_{x' \in \mathcal{X}} q_\beta(y|x') \left| \frac{\hat{q}_b(y|x')}{q_\beta(y|x')} - 1 \right| \left| \sum_{x \in \mathcal{X}} \hat{p}_a(x'|x) w_x \right| \\ \geq& 2^{-1} \tilde{\varepsilon}_\theta \sum_{x' \in \mathcal{X}} q_\beta(y|x') \\ \geq& 2^{-1} \tilde{\varepsilon}_\theta q_\beta(y|1) \\ \geq& 4^{-1} \tilde{\varepsilon}_\theta |\hat{q}_b(y|1)| \end{aligned} \tag{132}$$

for any $a \in V_{\delta_\theta}(\alpha)$, $b \in V_{\delta_\theta}(\beta)$, $\eta = [a^T b^T]^T$, $y \in [-\tilde{C}_\theta, \tilde{C}_\theta]$, $w = [w_1 \cdots w_{N_x}]^T \in V_{\delta_\theta}(\mathcal{P}^{N_x})$. Then, it can concluded from (131), (132) that $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are analytical in (η, w) on $V_{\delta_\theta}(\theta) \times V_{\delta_\theta}(\mathcal{P}^{N_x})$ for each $y \in \mathcal{Y}$. On the other side, (126), (130) – (132) imply

$$\begin{aligned} &|\hat{\phi}_\eta(w,y)| \leq |\log|e^T \hat{R}_\eta(y)w|| + 2\pi \leq \tilde{C}_\theta(1+y^2) + \log(N_x \tilde{C}_\theta \tilde{\varepsilon}_\theta^{-1}) + 2\pi \leq K_\theta(1+y^2), \\ &\|\hat{G}_\eta(w,y)\| \leq 4 N_x \tilde{C}_\theta \tilde{\varepsilon}_\theta^{-2} \leq K_\theta \end{aligned}$$

for any $\eta \in V_{\delta_\theta}(\theta)$, $w \in V_{\delta_\theta}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$. Hence, the lemma's assertion holds. ∎

Proof of Proposition 4: Let $Q \subset \Theta$ be an arbitrary compact set. Then, owing to Conditions (i), (ii) of the proposition, there exists a real number $\varepsilon_Q \in (0,1)$ such that $\varepsilon_Q \leq p_\alpha(x'|x) \leq \varepsilon_Q^{-1}$ for all $\alpha \in \mathcal{A}$, $x, x' \in \mathcal{X}$ satisfying $[\alpha^T \beta^T]^T \in Q$ for some $\beta \in \mathcal{B}$. Therefore,

$$\varepsilon_Q q_\beta(y|x') \leq r_\theta(y|x', x) \leq \varepsilon_Q^{-1} q_\beta(y|x')$$

for all $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$, $\theta = [\alpha^T \beta^T]^T$, $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ satisfying $\theta \in Q$. Thus, Assumption 3 is true.

Since the collection of sets $\{V_{\delta_\theta/2}(\theta)\}_{\theta \in Q}$ covers Q and since Q is compact, there exists a finite subset $\tilde{Q}$ of Q such that Q is covered by $\{V_{\delta_\theta/2}(\theta)\}_{\theta \in \tilde{Q}}$. Let $\delta_Q = \min\{\delta_\theta/2 : \theta \in \tilde{Q}\}$, $K_Q = \max\{K_\theta : \theta \in \tilde{Q}\}$ (δ_θ, K_θ are defined in the statement of Lemma 20). Obviously, $\delta_Q \in (0,1)$, $K_Q \in [1, \infty)$. It can also be deduced that for each $\theta \in Q$, $V_{\delta_Q}(\theta) \times V_{\delta_Q}(\mathcal{P}^{N_x})$ is contained in one of the sets from the collection $\{V_{\delta_\theta}(\theta)\}_{\theta \in \tilde{Q}}$. Thus, $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x}) \subseteq \bigcup_{\theta \in \tilde{Q}} V_{\delta_\theta}(\theta) \times V_{\delta_\theta}(\mathcal{P}^{N_x})$. Then, as an immediate consequence of Lemma 20, we have that Assumption 4 holds. ■

VI. Conclusion

We have studied the asymptotic properties of recursive maximum likelihood estimation in hidden Markov models. We have analyzed the asymptotic behavior of the asymptotic log-likelihood function and the convergence and convergence rate of the recursive maximum likelihood algorithm. Using the principle of analytical continuation, we have shown the analyticity of the asymptotic log-likelihood for analytically parameterized hidden Markov models. Relying on this result and Lojasiewicz inequality, we have demonstrated the point-convergence of the recursive maximum likelihood algorithm, and we have derived relatively tight bounds on the convergence rate. The obtained results cover a relatively broad class of hidden Markov models with finite state space and continuous observations. They can also be extended to batch (i.e., non-recursive) maximum likelihood estimators such as those studied in [6], [11], [24], [33]. In the future work, attention will be given to the possibility of extending the result of this paper to hidden Markov models with continuous state space. The possibility of obtaining similar rate of convergence results for non-analytically parameterized hidden Markov models will be explored, too.

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] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations,* Springer-Verlag, 1990.

[3] D. P. Bertsekas, *Nonlinear Programming,* 2nd edition, Athena Scientific, 1999.

[4] D. P. Bertsekas and J. N. Tsitsiklis, *Gradient convergence in gradient methods with errors,* SIAM Journal on Optimization, 10 (2000), pp. 627 – 642.

[5] P. J. Bickel and Y. Ritov, *Inference in Hidden Markov Models I. Local Asymptotic Normality in the Stationary Case,* Bernoulli, 2 (1996), pp. 199 – 228.

[6] P. J. Bickel, Y. Ritov, and T. Ryden, *Asymptotic Normality of the Maximum Likelihood Estimator for General Hidden Markov Models,* Annals of Statistics, 26 (1998), pp. 1614 – 1635.

[7] V. S. Borkar and S. P. Meyn, *The ODE Method for Convergence of Stochastic Approximation and Reinforcement Learning,* SIAM Journal on Control and Optimization, 38 (2000), pp. 447 – 469.

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

[9] H.-F. Chen, *Stochastic Approximation and Its Application,* Kluwer, 2002.

[10] R. Douc and C. Matias, *Asymptotics of the Maximum Likelihood Estimator for Hidden Markov Models for General Hidden Markov Models,* Bernoulli, 7 (2002), pp. 381 – 420.

[11] 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.

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

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

[14] 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.

[15] 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.

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

[17] K. Kurdyka, *On gradients of functions definable in o-minimal structures,* Annales de l'Institut Fourier (Grenoble), 48 (1998), pp. 769 - 783.

[18] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications,* 2nd edition, Springer-Verlag, 2003.

[19] F. Le Gland and L. Mével, *Recursive Identification of HMM's with Observations in a Finite Set,* Proceedings of the 34th Conference on Decision and Control, pp. 216 – 221, 1995.

[20] F. Le Gland and L. Mével, *Recursive Estimation in Hidden Markov Models,* Proceedings of the 36th Conference on Decision and Control, pp. 3468 – 3473, 1997.

[21] F. Le Gland and L. Mével, *Basic Properties of the Projective Product with Application to Products of Column-Allowable Nonnegative Matrices,* Mathematics of Control, Signals and Systems 13 (2000), pp 41 – 62.

[22] 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.

[23] 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.

[24] B. G. Leroux, *Maximum-Likelihood Estimation for Hidden Markov Models,* Stochastic Processes and Their Applications, 40 (1992), pp. 127 – 143.

[25] L. Ljung, *System Identification: Theory for the User,* 2nd edition, Prentice Hall, 1999.

[26] S. Lojasiewicz, *Sur le problème de la division,* Studia Mathematica, 18 (1959), pp. 87 – 136.

[27] S. Lojasiewicz, *Sur la géométrie semi- et sous-analytique,* Annales de l'Institut Fourier (Grenoble), 43 (1993), pp. 1575 – 1595.

[28] L. Mevel and L. Finesso, *Asymptotical Statistics of Misspecified Hidden Markov Models,* IEEE Transactions on Automatic Control, 49 (2004), pp. 1123 – 1132.

[29] 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.

[30] 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.

[31] B. T. Polyak and Y. Z. Tsypkin, *Criterion algorithms of stochastic optimization,* Automation and Remote Control, 45 (1984), pp. 766 – 774.

[32] B. T. Polyak, *Introduction to Optimization,* Optimization Software, 1987.

[33] T. Ryden, *Consistent and Asymptotically Normal Parameter Estimates for Hidden Markov Models,* Annals of Statistics, 22 (1994), pp. 1884 -1895.

[34] T. Ryden, *On recursive estimation for hidden Markov models,* Stochastic Processes and Their Applications 66 (1997), pp. 79 – 96.

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

[36] 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.

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