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

Vladislav B. Tadić

Abstract—**This paper considers the asymptotic properties of the recursive maximum-likelihood estimator for 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 analytic 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—**Analyticity, convergence rate, hidden Markov models, Lojasiewicz inequality, maximum-likelihood estimation, point convergence, recursive identification.**

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 the 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, computational 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], [13], and references cited therein).

Besides the estimation of states given available observations (also known as filtering), the identification of model parameters is 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 observation likelihoods given available observations. The identification of hidden Markov models has been considered in numerous papers and several methods and algorithms have been developed (see [8, Part II], [13], and references cited therein). Among them, the methods based on the maximum-likelihood principle are probably the most important. Their various asymptotic properties (asymptotic consistency, asymptotic normality, convergence rate) have been analyzed in a number of papers (see [1], [5], [6], [11], [12], [19], [22]–[25], [27], [31], [36], [37]; see also [8, ch. 12], [13], 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 closed-form expression and is fairly complex even for small-size hidden Markov models (four or more states), it is hard (if not impossible) 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 log-likelihood (mainly entropy rate) have recently been studied in [15]–[17], [32], [33], and [38]. 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 estimator for hidden Markov models with a discrete state-space and continuous observations. We establish a link between the analyticity of the asymptotic log-likelihood on the one hand, and the point convergence and convergence rate of the recursive maximum-likelihood algorithm, on the other hand. More specifically, relying on the principle of analytic continuation, we show under mild conditions that the asymptotic log-likelihood function is analytic 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

Manuscript received May 18, 2009; revised February 14, 2010. Date of current version November 19, 2010.

The author is with the Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K. (e-mail: v.b.tadic@bristol.ac.uk).

Communicated by F. Hlawatsch, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2010.2081110

without assuming that the log-likelihood function has an isolated strong maximum.

The paper is organized as follows. In Section II, hidden Markov models and the corresponding recursive maximum-likelihood algorithms are defined. The main results are also presented in Section II. 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 proved 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 in $\mathbb{R}^{d_y}$. $\{p(x'|x)\}_{x,x' \in \mathcal{X}}$ are nonnegative 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 in $\mathcal{Y}$. On the other side, d_θ is a positive integer, while Θ is an open set in $\mathbb{R}^{d_\theta}$. $\{p_\theta(x'|x)\}_{x,x' \in \mathcal{X}}$ 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

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

w.p.1 for each $x \in \mathcal{X}$, $n \geq 0$ and any Borel-measurable set B in $\mathcal{Y}$. Finally, $f(\cdot)$ stands for the asymptotic log-likelihood 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

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

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 the (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 asymptotic likelihood $f(\cdot)$ 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], [21], [35], 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 $e = [1 \dots 1]^T \in \mathbb{R}^{N_x}$. With this notation, a stochastic gradient algorithm for maximizing $f(\cdot)$ can be defined as

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

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

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

In this recursion, $\{\alpha_n\}_{n \geq 0}$ is 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 are 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 (3) usually includes a projection (or truncation) device which prevents estimates $\{\theta_n\}_{n \geq 0}$ from leaving Θ (see [9] and [28] for further details). As the problems studied in the paper are already complex, this aspect of algorithm (3) is not considered here. Instead, similarly as in [2, Part II], [21], and [28], our results on the asymptotic behavior of algorithm (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.,

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

Sequence $\{\gamma_n\}_{n\geq 0}$ is defined by $\gamma_0 = 1$ and

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

for $n \geq 1$.

Algorithm (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 irreducible and aperiodic (i.e., 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) analytic 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 analytic 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

$$\begin{aligned} |\hat{\phi}_\eta(w, y)| &\leq \psi_Q(y) \\ \|\hat{G}_\eta(w, y)\| &\leq K_Q \\ \int \psi_Q^2(y') Q(dy' \,|\, x) &< \infty \end{aligned} \tag{4}$$

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], [11], [12], [22]–[25], [27], [31], [36], [37], and [39]; 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 conditional densities $q_\theta(y \,|\, x)$ to be analytic in θ. It also requires $p_\theta(x' \,|\, x)$ and $q_\theta(y \,|\, x)$ to be analytically continuable to a complex domain in such a way 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.

The main purpose of Assumption 4 is to ensure that the optimal filter associated with the transfer function $G_\theta(u, y)$ is analytically continuable to a complex domain (see Lemma 4). Since the asymptotic log-likelihood $f(\cdot)$ can be represented as a limit of this filter, Assumption 4 (together with the limit theorems for complex-analytic functions) also ensures the analyticity of $f(\cdot)$ (see Theorem 1 and its proof). On the other side, the asymptotic behavior of algorithm (3) (point convergence and convergence rate) crucially relies on this property of $f(\cdot)$ (see Theorems 2 and 3 and their proofs; see also the outline of the proofs provided in Section IV-A).

In order to state our main results, we rely on the following notation. 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 the asymptotic likelihood $f(\cdot)$ and algorithm (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

$$|f(\theta') - f(\theta)| \leq M_\theta \|\nabla f(\theta')\|^{\mu_\theta} \tag{5}$$

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

Theorem 2 (Convergence): Let Assumptions 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

$$\|\nabla f(\theta_n)\|^2 = O\left(\gamma_n^{-\hat{p}}\right) \tag{6}$$

$$|f(\theta_n) - f(\hat{\theta})| = O\left(\gamma_n^{-\hat{p}}\right) \tag{7}$$

$$\|\theta_n - \hat{\theta}\| = O\left(\gamma_n^{-\hat{q}}\right) \tag{8}$$

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

$$\hat{r} = \begin{cases} 1/(2 - \hat{\mu}), & \text{if } \hat{\mu} < 2 \\ \infty, & \text{otherwise} \end{cases} \tag{9}$$

$$\hat{p} = \hat{\mu} \min\{r, \hat{r}\} \tag{10}$$

$$\hat{q} = \min\{r, \hat{r}\} - 1. \tag{11}$$

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

In the literature on deterministic and stochastic optimization, the convergence of gradient search is usually characterized 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], [34], [35], 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 (3), this kind of information is provided by Theorems 2 and 3. Basically, Theorem 2 claims that recursion (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 the convergence rate in the terms of Lojasiewicz exponent $\hat{\mu} = \mu_{\hat{\theta}}$ 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 (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., [2], [21], [28], and references cited therein).

Various asymptotic properties of maximum-likelihood estimation in hidden Markov models have been analyzed thoroughly in a number of papers [1], [5], [6], [11], [12], [22]–[25], [27], [31], [36], [37]; (see also [8, ch. 12], [13], 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 the objective function to have an isolated maximum at which the Hessian is strictly negative definite. Since $f(\cdot)$, the objective function of recursion (3), is rather complex even when the observation space is finite (i.e., $\mathcal{Y} = \{1,\ldots,N_y\}$) 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 nonisolated maximum or minimum). In addition to this, the theorems cover a relatively broad class of hidden Markov models (see Section III). 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 [15]–[17], [32], [33], [38]. 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 the 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 analytic continuation and is similar to the methodology formulated in [15].

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 section. For these models, we show that the conclusions of Theorems 1–3 hold whenever the parameterization of candidate models is analytic.

Let $N_y > 2$ be an integer, while $\mathcal{Y} = \{1,\ldots,N_y\}$. 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 analytic 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 and 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.

[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

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

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

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

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

B. Compactly Supported Observations

In this section, 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 in $\mathbb{R}^{d_y}$. 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 analytic 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 and 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 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 latter 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.

C. Mixture of Observation Likelihoods

In this section, 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

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

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 section, 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 analytic 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 and 2 and the conditions of Proposition 4 hold. Then, the conclusions of Theorems 1–3 are true.

The proof is provided in Section V.

D. Gaussian Observations

This section is devoted to hidden Markov models with a finite number of states and with Gaussian observations. More specifically, d_α and $\mathcal{A}$ have the same meaning as in the previous section, while $\mathcal{Y} = \mathbb{R}$ and $\mathcal{B} = ((0,\infty)^{N_x} \times \mathbb{R}^{N_x}) \setminus \tilde{\mathcal{B}}$, where

$$\tilde{\mathcal{B}} = \left\{ [\lambda_1 \cdots \lambda_{N_x} \mu_1 \cdots \mu_{N_x}]^T \in (0,\infty)^{N_x} \times \mathbb{R}^{N_x} : \lambda_{x'} = \lambda_{x''} = \min_{x \in \mathcal{X}} \lambda_x \text{ for some } x' \neq x'', x', x'' \in \mathcal{X} \right\}. \tag{12}$$

Similarly, as in Section III-C, 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 section, 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 analytic 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 and 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 Section III-D do not satisfy Assumption 4 without the condition that $\arg\min_{x \in \mathcal{X}} \lambda_x$ is a singleton (the details are provided in Appendix III). 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. Outline of the Proof

Theorems 1–3 are proved in several stages. The proofs are presented in Sections IV-B–IV-E. The main steps can be summarized as follows.

Section IV-B is mainly focused on the stability properties of optimal filter $G_{\theta,\boldsymbol{y}}^{0:n}(u)$ and its analytic continuation $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ (to be defined in Section IV-D). It is also concerned with the stability of filter derivatives $H_{\theta,\boldsymbol{y}}^{0:n}(u, V)$ (also to be defined in Section IV-B) and with the analytical properties of functions $F_\theta(u, V, y)$ and $\hat{G}_\eta(w, y)$. In Lemma 1, the analytical properties (local boundedness and Lipschitz continuity) of $F_\theta(u, V, y)$

and $\hat{G}_\eta(w,y)$ are studied, while Lemmas 2 and 3 consider the stability properties (forgetting, boundedness, and ergodicity) of $G_{\theta,\boldsymbol{y}}^{0:n}(u)$, $H_{\theta,\boldsymbol{y}}^{0:n}(u,V)$. Lemma 1 is a consequence of Assumption 4 and the Cauchy inequality for complex-valued analytic functions, while Lemmas 2 and 3 are based on the existing results on the stability of the optimal filter. Lemmas 1–3 are necessary prerequisite for Lemmas 4 and 5 and Theorem 1. The most important results of Section IV-B are contained in Lemma 4. The lemma deals with the stability of the analytic continuation $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ of the optimal filter. Starting with the results of Lemma 2 and relying on the principle of analytic continuation, the lemma shows that $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ forgets initial condition w geometrically in n and uniformly in η, w, $\boldsymbol{y}$. This result is the foundation of Theorem 1 and Corollary 5.

Section IV-C is focused on the properties of the asymptotic log-likelihood $f(\theta)$. In this section, the analyticity of $f(\theta)$ is proved (Theorem 1) and the most general version of the Lojasiewicz inequality for $f(\theta)$ is provided (Corollary 5). These results are a crucial prerequisite for the asymptotic analysis of algorithm (3) (carried out in Section IV-E) and Theorems 2 and 3. In particular, without Lojasiewicz inequality (32) (notice that the analyticity is required by any version of this inequality), it is not possible to establish the results of Lemma 10 [i.e., inequalities (59) and (60)], which themselves almost directly lead to the convergence rate of $\{f(\theta_n)\}_{n\geq 0}$ and $\{\nabla f(\theta_n)\}_{n\geq 0}$ (Lemmas 11–13). The proof of Theorem 1 is based on the principle of analytic continuation and the fact that the limit of uniformly convergent complex-analytic functions is also analytic. Since this fact is not true for real-analytic functions (and since $f(\theta)$ is a function of the optimal filter), it is necessary to demonstrate the geometric forgetting not only for $G_{\theta,\boldsymbol{y}}^{0:n}(u)$, but also for its analytic continuation $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ (Lemma 4).

Section IV-D provides an equivalent representation of algorithm (3) and studies its immediate properties. More specifically, the section shows that recursion (3) is a stochastic gradient search with additive noise. The section also provides the basic asymptotic properties of the noise sequence $\{\xi_n\}_{n\geq 0}$ (to be defined in Section IV-D) and sequences $\{f(\theta_n)\}_{n\geq 0}$, $\{\nabla f(\theta_n)\}_{n\geq 0}$. In Lemma 5, the Poisson (34) associated with algorithm (3) is analyzed and the basic properties of its solution are demonstrated. Lemmas 6 and 7 study the asymptotic behavior of $\{\xi_n\}_{n\geq 0}$, $\{f(\theta_n)\}_{n\geq 0}$ and $\{\nabla f(\theta_n)\}_{n\geq 0}$: in Lemma 6, an upper bound on the convergence rate of $\{\xi_n\}_{n\geq 0}$ is derived, while the convergence of $\{f(\theta_n)\}_{n\geq 0}$, $\{\nabla f(\theta_n)\}_{n\geq 0}$ is proved in Lemma 7. Lemma 5 follows from Lemmas 1–3 and the results of [2, ch. II.2], while Lemma 6 is based on Lemma 5 and the techniques developed in [2, ch. II.1]. Lemmas 6 and 7 are important prerequisite for the asymptotic analysis conducted in Section IV-E, i.e., for Lemma 8 and the construction of Lyapunov functions $u(\theta)$, $v(\theta)$ (notice that both $u(\theta)$ and $v(\theta)$ depend on $\hat{f} = \lim_{n\to\infty} f(\theta_n)$). In Section IV-E, the asymptotic behavior of algorithm (3) is analyzed and Theorems 2 and 3 are proved. The main steps in the analysis can be summarized as follows.

Step 1: Further asymptotic properties of $\{\xi_n\}_{n\geq 0}$, $\{f(\theta_n)\}_{n\geq 0}$, and $\{\nabla f(\theta_n)\}_{n\geq 0}$ are provided in Lemmas 8–10. These lemmas are based on Taylor formula for Lyapunov functions $u(\theta)$, $v(\theta)$ and Bellman–Gronwall and Lojasiewicz inequalities (50) and (42). It is important to emphasize that the standard form of Lojasiewicz inequality (5) (provided in Theorem 1) cannot be applied at this stage of the analysis due to the following reasons: i) (5) holds only locally in a close vicinity of $\theta \in \Theta$, and ii) there are no guarantees that $\inf_{\theta\in Q}\delta_\theta > 0$, $\sup_{\theta\in Q} M_\theta < \infty$ for any compact set $Q \subset \Theta$ (δ_θ and M_θ are specified in Theorem 1). Hence, without knowing that $\lim_{n\to\infty}\theta_n = \hat{\theta}$ exists (which becomes evident in Lemma 15, at the very end of the analysis), it is not possible to use (5) to analyze the asymptotic behavior of $\{\theta_n\}_{n\geq 0}$. As opposed to (5), inequality (42) (which is a direct consequence of the form of Lojasiewicz inequality provided in Corollary 5) can be applied to the asymptotic analysis of $\{\theta_n\}_{n\geq 0}$. The reasons are as follows: i) $\lim_{n\to\infty} f(\theta_n) = \hat{f}$ exists (due to Lemma 7), and ii) (42) is satisfied by all $\theta \in \hat{Q}$ for which $f(\theta)$ is sufficiently close to $\hat{f}$ ($\hat{Q}$ is defined in the beginning of Section IV-E and represents a compact set whose interior contains the limit point of $\{\theta_n\}_{n\geq 0}$).

Step 2: Relying on the results of Step 1 (Lemma 10), $\limsup_{n\to\infty}\gamma_n^{\hat{p}}(f(\theta_n) - \hat{f}) < \infty$ is proved in Lemma 11. The idea of the proof can be described as follows. If the previous relation is not true, then there exists an increasing sequence $\{n_k\}_{k\geq 0}$ such that $\lim_{k\to\infty}\gamma_{n_k}^{\hat{p}} u(\theta_{n_k}) = -\infty$. Consequently, the Lojasiewicz inequality (42) implies $\lim_{k\to\infty}\gamma_{n_k}^{r}\|\nabla f(\theta_{n_k})\| = \infty$ (notice that $\hat{p}/\hat{\mu} \leq r$). Then, the Taylor formula for $u(\theta)$ and the algorithm's representation (33) yield

$$\begin{aligned}
&u(\theta_n)\\
&\approx u\left(\theta_{n_k}\right) - \left(\nabla f\left(\theta_{n_k}\right)\right)^T \sum_{i=n_k}^{n-1}\alpha_i(\nabla f(\theta_i) + \xi_i)\\
&\approx -\left(\nabla f\left(\theta_{n_k}\right)\right)^T\left(\left(\gamma_n - \gamma_{n_k}\right)\nabla f\left(\theta_{n_k}\right) + \sum_{i=n_k}^{n-1}\alpha_i\xi_i\right)\\
&\quad + u\left(\theta_{n_k}\right)\\
&\leq -\left\|\nabla f\left(\theta_{n_k}\right)\right\|\left(\left(\gamma_n - \gamma_{n_k}\right)\left\|\nabla f\left(\theta_{n_k}\right)\right\| - \left\|\sum_{i=n_k}^{n-1}\alpha_i\xi_i\right\|\right)\\
&\quad + u\left(\theta_{n_k}\right)
\end{aligned}$$

for $n \geq n_k$ and all sufficiently large $k \geq 0$. As $\gamma_n - \gamma_{n_k} \geq 1$ for $n > a(n_k, 1)$ and $\max_{n\geq n_k}\left\|\sum_{i=n_k}^{n}\alpha_i\xi_i\right\| = o(\gamma_{n_k}^{-r})$ (due to Lemma 8), we get $u(\theta_n) \leq u(\theta_{n_k}) < 0$ for $n > a(n_k, 1)$ and all sufficiently large $k \geq 0$. However, this is not possible, since $\lim_{n\to\infty} u(\theta_n) = 0$ (due to Lemma 7).

Step 3: $\liminf_{n\to\infty}\gamma_n^{\hat{p}}(f(\theta_n) - \hat{f}) > -\infty$ is proved in Lemmas 12 and 13. The idea of the proof can be summarized as follows. If the previous relation is not satisfied, then there exists an increasing sequence $\{n_k\}_{k\geq 0}$ such that $\lim_{k\to\infty}\gamma_{n_k}^{\hat{p}} u(\theta_{n_k}) = \infty$. Consequently, $\lim_{k\to\infty}\gamma_{n_k}^{-1} v(\theta_{n_k}) = 0$, while the Lojasiewicz inequality (42) yields $\lim_{k\to\infty}\gamma_{n_k}^{r}\|\nabla f(\theta_{n_k})\| = \infty$. The same inequality also implies

$$\begin{aligned}
\left\|\nabla f\left(\theta_{n_k}\right)\right\|^2 &\geq \hat{M}^{-2/\hat{\mu}}\left(\hat{f} - f\left(\theta_{n_k}\right)\right)^{2/\hat{\mu}}\\
&\geq 2\hat{p}\hat{L}(u(\theta_{n_k}))^{1+1/\hat{p}}
\end{aligned}$$

for all sufficiently large $k \geq 0$ and $\hat{L} = 2^{-1}\hat{p}^{-1}\hat{M}^{-2/\hat{\mu}}$ (notice that $2/\hat{\mu} = 1 + 1/(\hat{\mu}\hat{r}) \leq 1 + 1/\hat{p}$ and that $u(\theta_{n_k}) \approx 0$ for sufficiently large k). Then, owing to the Taylor formula for $v(\theta)$ and the algorithm's representation (33), we have

$$\begin{aligned}
&v(\theta_n)\\
&\approx v(\theta_{n_k}) + \frac{(\nabla f(\theta_{n_k}))^T}{\hat{p}(u(\theta_{n_k}))^{1+1/\hat{p}}} \sum_{i=n_k}^{n-1} \alpha_i (\nabla f(\theta_i) + \xi_i)\\
&\approx \frac{(\nabla f(\theta_{n_k}))^T}{\hat{p}(u(\theta_{n_k}))^{1+1/\hat{p}}} \left((\gamma_n - \gamma_{n_k}) \nabla f(\theta_{n_k}) + \sum_{i=n_k}^{n-1} \alpha_i \xi_i \right)\\
&\quad + v(\theta_{n_k})\\
&\geq \frac{\|\nabla f(\theta_{n_k})\|}{\hat{p}(u(\theta_{n_k}))^{1+1/\hat{p}}} \left(\frac{(\gamma_n - \gamma_{n_k}) \|\nabla f(\theta_{n_k})\|}{2} - \left\| \sum_{i=n_k}^{n-1} \alpha_i \xi_i \right\| \right)\\
&\quad + \frac{(\gamma_n - \gamma_{n_k}) \|\nabla f(\theta_{n_k})\|^2}{2\hat{p}(u(\theta_{n_k}))^{1+1/\hat{p}}} + v(\theta_{n_k})\\
&\geq v(\theta_{n_k}) + \hat{L}(\gamma_n - \gamma_{n_k})
\end{aligned}$$

for $n > a(n_k, 1)$ and all sufficiently large $k \geq 0$ (notice that $\gamma_n - \gamma_{n_k} \geq 1$ for $n > a(n_k, 1)$ and that $\max_{n \geq n_k} \left\| \sum_{i=n_k}^{n} \alpha_i \xi_i \right\| = o(\gamma_{n_k}^{-r})$ follows from Lemma 8). Therefore, $\liminf_{n\to\infty} \gamma_n^{-1} v(\theta_n) \geq \hat{L} > 0$, which contradicts $\lim_{k\to\infty} \gamma_{n_k}^{-1} v(\theta_{n_k}) = 0$.

Step 4: $\|\nabla f(\theta_n)\|^2 = O(\gamma_n^{-\hat{p}})$ is shown (a direct consequence of Lemmas 11 and 13). The proof is based on the following reasoning. Due to the Taylor formula for $u(\theta)$ and the algorithm's representation (33), we have

$$\begin{aligned}
&\|\nabla f(\theta_n)\|^2\\
&\quad\approx \frac{u(\theta_n) - u(\theta_k)}{\gamma_k - \gamma_n} - \frac{(\nabla f(\theta_n))^T}{\gamma_k - \gamma_n} \sum_{i=n}^{k-1} \alpha_i \xi_i\\
&\quad\leq \frac{1}{2(\gamma_k - \gamma_n)} \left(\|\nabla f(\theta_n)\|^2 + \left\| \sum_{i=n}^{k-1} \alpha_i \xi_i \right\|^2 \right)\\
&\qquad + \frac{|u(\theta_k)| + |u(\theta_n)|}{\gamma_k - \gamma_n}
\end{aligned}$$

for $k \geq n$ and all sufficiently large $n \geq 0$. Consequently, setting $k = a(n, 1)$ and using the results of Steps 2 and 3, we get

$$\begin{aligned}
\|\nabla f(\theta_n)\|^2 &\leq 2\left|u\left(\theta_{a(n,1)}\right)\right| + 2|u(\theta_n)| + \left\| \sum_{i=n}^{a(n,1)-1} \alpha_i \xi_i \right\|^2\\
&= O\left(\gamma_n^{-\hat{p}}\right)
\end{aligned}$$

for all sufficiently large $n \geq 0$ (notice that $\gamma_{a(n,1)} - \gamma_n \approx 1$).

Step 5: $\max_{k \geq n} \|\theta_k - \theta_n\| = O(\gamma_n^{-\hat{q}})$ is demonstrated in Lemmas 14 and 15. The proof is based on the results of Steps 2–4 and the following relations:

$$\begin{aligned}
\|\theta_k - \theta_n\| &\approx \left\| (\gamma_k - \gamma_n) \nabla f(\theta_n) + \sum_{i=n}^{k-1} \alpha_i \xi_i \right\|\\
&\leq (\gamma_k - \gamma_n) \|\nabla f(\theta_n)\| + \left\| \sum_{i=n}^{k-1} \alpha_i \xi_i \right\|\\
&\approx \frac{1}{\|\nabla f(\theta_n)\|} \left(u(\theta_n) - u(\theta_k) - (\nabla f(\theta_n))^T \sum_{i=n}^{k-1} \alpha_i \xi_i \right)\\
&\quad + \left\| \sum_{i=n}^{k-1} \alpha_i \xi_i \right\|\\
&\leq \frac{u(\theta_n) - u(\theta_k)}{\|\nabla f(\theta_n)\|} + 2\left\| \sum_{i=n}^{k-1} \alpha_i \xi_i \right\|
\end{aligned}$$

where $k \geq n$ and $n \geq 0$ is sufficiently large.

Step 6: Theorems 2 and 3 are proved. The convergence and convergence rate of $\{\theta_n\}_{n\geq 0}$ directly follow from the results of Step 5, while the convergence rates of $\{f(\theta_n)\}_{n\geq 0}$, $\{\nabla f(\theta_n)\}_{n\geq 0}$ are immediate consequences of Steps 2–4. As $\lim_{n\to\infty} \theta_n = \hat{\theta}$, $\hat{p}$ and $\hat{q}$ can be defined by (10) and (11), and $\hat{\mu} = \mu_{\hat{\theta}}$, the Lojasiewicz exponent at $\hat{\theta}$ (notice that in Lemmas 10–15 and their proofs, the definition of $\hat{p}$ and $\hat{q}$ is based on $\hat{\mu} = \mu_{\hat{Q},\hat{f}}$, another Lojasiewicz exponent specified in Corollary 5; also notice that the tighter convergence rates are obtained with $\hat{\mu} = \mu_{\hat{\theta}}$ than with $\hat{\mu} = \mu_{\hat{Q},\hat{f}}$).

B. Optimal Filter and Its Properties

The stability properties (forgetting and ergodicity) of the optimal filter (2), its derivatives (3), and its analytic continuation (to be defined in the next paragraph) are studied in this section. The analytical properties (local boundedness and local Lipschitz continuity) of functions $F_\theta(u, V, y)$ and $\hat{G}_\eta(w, y)$ are also considered here. The analysis mainly follows the ideas and results of [24], [25], and [26].

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

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

($G_\theta(u, y)$, $\hat{G}_\eta(w, y)$, and $H_\theta(u, V, y)$ are defined in Section II). If $\boldsymbol{\vartheta} = \{\theta\}_{n\geq 0}$ (i.e., $\vartheta_n = \theta$), we also use notation $G_{\theta,\boldsymbol{y}}^{m:n}(u) = G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u)$, $H_{\theta,\boldsymbol{y}}^{m:n}(u, V) = H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u, V)$. 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 $\hat{G}_\eta^{0:n}(w, y_{1:n}) = \hat{G}_{\boldsymbol{\eta},\boldsymbol{y}}^{0:n}(w)$.

In this section, 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 $\Pi_\theta(\cdot,\cdot)$ are the transition kernels of Markov chains

$$\left\{X_{n+1}, Y_{n+1}, G_{\theta,\boldsymbol{Y}}^{0:n}(u), H_{\theta,\boldsymbol{Y}}^{0:n}(u,V)\right\}_{n\geq 0}$$
$$\left\{X_{n+1}, Y_{n+1}, G_{\theta,\boldsymbol{Y}}^{0:n}(u)\right\}_{n\geq 0}$$

where $\boldsymbol{Y} = \{Y_n\}_{n\geq 0}$ (notice that $P_\theta(\cdot,\cdot)$, $\Pi_\theta(\cdot,\cdot)$ 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

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

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

$$\|F(\theta,z)\| \leq C_{1,Q}\psi_Q(y)(1+\|V\|)$$
$$\|F(\theta',z) - F(\theta'',z)\| \leq C_{1,Q}\psi_Q(y)\|\theta'-\theta''\|$$
$$\|\hat{G}_{\eta'}(w',y) - \hat{G}_{\eta''}(w'',y)\| \leq C_{1,Q}(\|\eta'-\eta''\| + \|w'-w''\|)$$

for all $\theta,\theta',\theta'' \in Q$, $\eta',\eta'' \in V_{\delta_{1,Q}}(Q)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, and $z = (x,y,u,V)$.

Remark: Lemma 1 is a special case of Lemma 17 (provided in Appendix I) and a direct consequence of Assumption 4 and the Cauchy inequality for complex analytic functions.

Lemma 2: Let Assumptions 3 and 4 hold. Then, for any compact set $Q \subset \Theta$, there exist real numbers $\varepsilon_{1,Q} \in (0,1)$, $C_{2,Q} \in [1,\infty)$ such that

$$\left\|G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(w') - G_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(w'')\right\| \leq C_{2,Q}\varepsilon_{1,Q}^{n-m}\|w'-w''\| \quad (13)$$
$$\left\|H_{\boldsymbol{\vartheta},\boldsymbol{y}}^{m:n}(u,V)\right\| \leq C_{2,Q}(1+\|V\|) \quad (14)$$

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

Remark: Lemma 2 is an extension of the results of [10], [24]–[26], and [39]. It is proved in Appendix I.

Lemma 3: Let Assumptions 2–4 hold. Then, $f(\cdot)$ is well defined and differentiable on Θ. Moreover, for any compact set $Q \subset \Theta$, there exist real numbers $\varepsilon_{2,Q} \in (0,1)$, $C_{3,Q} \in [1,\infty)$ such that

$$|(\Pi^n\phi)(\theta,\zeta) - f(\theta)| \leq C_{3,Q}\varepsilon_{2,Q}^n \quad (15)$$
$$\|(P^nF)(\theta,z) - \nabla f(\theta)\| \leq C_{3,Q}\varepsilon_{2,Q}^n(1+\|V\|^2) \quad (16)$$
$$\|(P^nF)(\theta',z) - (P^nF)(\theta'',z)\| \leq C_{3,Q}\varepsilon_{2,Q}^n\|\theta'-\theta''\|(1+\|V\|^2) \quad (17)$$

for all $\theta,\theta',\theta'' \in Q$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $z \in (x,y,u,V) \in \mathcal{S}_z$, $n \geq 0$, where

$$(\Pi^n\phi)(\theta,\zeta) = \int \phi(\theta,\zeta')\Pi_\theta^n(\zeta,d\zeta')$$
$$(P^nF)(\theta,z) = \int F(\theta,z')P_\theta^n(z,dz').$$

Remark: Lemma 3 is an extension of the results of [24], [25], and [39]. Its proof is provided in Appendix I.

Lemma 3: Let Assumptions 3 and 4 hold. Then, for any compact set $Q \subset \Theta$, there exist real numbers $\delta_{2,Q}, \varepsilon_{3,Q} \in (0,1)$, $C_{4,Q} \in [1,\infty)$ such that the following is true.

i) $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ is analytic in (η,w) on $V_{\delta_{2,Q}}(Q) \times V_{\delta_{2,Q}}(\mathcal{P}^{N_x})$ for each $n \geq 0$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in $\mathcal{Y}$.

ii) Inequalities

$$d(\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w), \mathcal{P}^{N_x}) \leq \min\{\delta_Q, \delta_{1,Q}\}$$
$$\left\|\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w') - \hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w'')\right\| \leq C_{4,Q}\varepsilon_{3,Q}^n\|w'-w''\|$$

hold for all $\eta \in V_{\delta_{2,Q}}(Q)$, $w,w',w'' \in V_{\delta_{2,Q}}(\mathcal{P}^{N_x})$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in $\mathcal{Y}$ (δ_Q is specified in Assumption 4).

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

First, we prove by induction (in k) that

$$d\left(\hat{G}_{\eta,\boldsymbol{y}}^{n:n+k}(w), \mathcal{P}^{N_x}\right) \leq \left(2^{k+1}C_{1,Q}^k - 1\right)\tilde{\delta}_{2,Q} \leq \tilde{\delta}_{1,Q} \quad (18)$$

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, (18) 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 (18) 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 1 implies

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

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

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

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, (18) 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 $\hat{G}_{\eta,\boldsymbol{y}}^{n:n}(w) = w$ and $\hat{G}_{\eta,\boldsymbol{y}}^{n:n+k+1}(w) = \hat{G}_\eta(\hat{G}_{\eta,\boldsymbol{y}}^{n:n+k}(w), y_{n+k+1})$, it can be deduced from Assumption 4 and (18) that $\hat{G}_{\eta,\boldsymbol{y}}^{n:n+k}(w)$ 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 as well). Due to Assumption 4 and (18), we also have

$$\left\|\hat{G}_{\eta,\boldsymbol{y}}^{n:n+k+1}(w)\right\| = \left\|\hat{G}_\eta\left(\hat{G}_{\eta,\boldsymbol{y}}^{n:n+k}(w), y_{n+k+1}\right)\right\| \leq K_Q \quad (19)$$

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 (19), 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 $\tilde{C}_{1,Q} = 4(d_\theta + N_x)K_Q/\tilde{\delta}_{2,Q}^2$) such that

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

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

$$\begin{aligned} \max\Big\{ &\left\| \hat{G}_{\eta',\boldsymbol{y}}^{n:n+k}(w') - \hat{G}_{\eta'',\boldsymbol{y}}^{n:n+k}(w'') \right\|, \\ &\left\| \nabla_w \hat{G}_{\eta',\boldsymbol{y}}^{n:n+k}(w') - \nabla_w \hat{G}_{\eta'',\boldsymbol{y}}^{n:n+k}(w'') \right\| \Big\} \\ &\leq \tilde{C}_{2,Q}(\|\eta' - \eta''\| + \|w' - w''\|) \end{aligned} \tag{20}$$

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

$$\begin{aligned} &\left\| G_{\theta,\boldsymbol{y}}^{n:n+k_Q}(u') - G_{\theta,\boldsymbol{y}}^{n:n+k_Q}(u'') \right\| \\ &\quad \leq C_{2,Q}\varepsilon_{1,Q}^{k_Q}\|u' - u''\| \leq (\varepsilon_{1,Q}/2)\|u' - u''\| \end{aligned}$$

for all $\theta \in Q$, $u', u'' \in \mathcal{Q}^{N_x}$, $n \geq 0$. Therefore, $\|\nabla_u G_{\theta,\boldsymbol{y}}^{n:n+k_Q}(w)\| \leq \varepsilon_{1,Q}/2$ for each $\theta \in Q$, $u \in \mathcal{P}^{N_x}$, $n \geq 0$, which, together with (20) yields

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

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

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

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

$$\begin{aligned} &\left\| \hat{G}_{\eta,\boldsymbol{y}}^{n:n+k_Q}(w') - \hat{G}_{\eta,\boldsymbol{y}}^{n:n+k_Q}(w'') \right\| \\ &\quad \leq \int_0^1 \left\| \nabla_w \hat{G}_{\eta,\boldsymbol{y}}^{n:n+k_Q}(tw' + (1-t)w'') \right\| \|w' - w''\| dt \\ &\quad \leq \varepsilon_{1,Q}\|w' - w''\| \end{aligned} \tag{21}$$

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

$$d\left(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w), \mathcal{P}^{N_x}\right) \leq \tilde{\delta}_{4,Q} \tag{22}$$

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, (22) 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 (22) 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, (20) and (21) imply

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

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

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

for each $\eta \in V_{\tilde{\delta}_{5,Q}}(Q)$, $w \in V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$. Hence, (22) 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 $\delta_{2,Q} = \min\{\tilde{\delta}_{4,Q}, \tilde{\delta}_{5,Q}\}$. As $\hat{G}_{\eta,\boldsymbol{y}}^{0:0}(w) = w$ and $\hat{G}_{\eta,\boldsymbol{y}}^{0:(i+1)k_Q}(w) = \hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:(i+1)k_Q}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w))$, it can be deduced from (22) that $\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w)$ is analytic 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 $\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:(i+1)k_Q}(w)$ 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 $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w) = \hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:n}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w))$ for $i = \lfloor n/k_Q \rfloor$, we conclude from (22) that $\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w)$ is analytic 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 analytic in (η, w) on $V_{\tilde{\delta}_{5,Q}}(Q) \times V_{\tilde{\delta}_{4,Q}}(\mathcal{P}^{N_x})$ for any $i \geq 0$, $0 \leq j \leq k_Q$). On the other side, (18) and (22) yield

$$\begin{aligned} d\left(\hat{G}_{\eta,\boldsymbol{y}}^{0:n}(w), \mathcal{P}^{N_x}\right) &= d\left(\hat{G}_{\eta,\boldsymbol{y}}^{ik_Q:n}(\hat{G}_{\eta,\boldsymbol{y}}^{0:ik_Q}(w)), \mathcal{P}^{N_x}\right) \\ &\leq \tilde{\delta}_{1,Q} = \min\{\delta_Q, \delta_{1,Q}\} \end{aligned} \tag{23}$$

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 $\varepsilon_{3,Q} = \varepsilon_{1,Q}^{1/k_Q}$, $C_{4,Q} = \tilde{C}_{2,Q}\varepsilon_{1,Q}^{-1}$. Owing to (21) and (22), we have

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

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

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

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, (20) and (22) yield

$$\begin{aligned}
&\left\| \hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w') - \hat{G}^{0:n}_{\eta,\boldsymbol{y}}(w'') \right\| \\
&\quad = \left\| \hat{G}^{ik_Q:n}_{\eta,\boldsymbol{y}}\left(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w')\right) - \hat{G}^{ik_Q:n}_{\eta,\boldsymbol{y}}\left(\hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w'')\right) \right\| \\
&\quad \leq \tilde{C}_{2,Q} \left\| \hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w') - \hat{G}^{0:ik_Q}_{\eta,\boldsymbol{y}}(w'') \right\| \\
&\quad \leq \tilde{C}_{2,Q}\varepsilon^i_{1,Q}\|w' - w''\| \\
&\quad \leq C_{4,Q}\varepsilon^n_{4,Q}\|w' - w''\|
\end{aligned}$$

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

C. Analyticity

In this section, using the results of the Section IV-B (Lemma 4), 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 [15].

Proof of Theorem 1: Let

$$\hat{\phi}^{m:n}_{\eta,\boldsymbol{y}}(w) = \hat{\phi}_\eta\left(\hat{G}^{m:n-1}_{\eta,\boldsymbol{y}}(w), y_n\right)$$

for $\eta \in \mathbb{C}^{d_\theta}$, $w \in \mathbb{C}^{N_x}$, $0 \leq m < n$ and any sequence $\{y_n\}_{n\geq 0}$ in $\mathcal{Y}$. Moreover, let $\boldsymbol{Y} = \{Y_n\}_{n\geq 0}$, while

$$\hat{\psi}^n_\eta(w,x) = E\left(\hat{\phi}^{0:n}_{\eta,\boldsymbol{Y}}(w)\,\middle|\, X_1 = x\right)$$

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

$$\begin{aligned}
&\hat{\psi}^{n+1}_\eta(w,x) \\
&\quad = E\left(E\left(\hat{\phi}^{1:n+1}_{\eta,\boldsymbol{Y}}\left(\hat{G}_\eta(w,Y_1)\right)\middle| X_1, X_2, Y_1\right)\middle| X_1 = x\right) \\
&\quad = E(\hat{\psi}^n_\eta(\hat{G}_\eta(w,Y_1), X_2) \,|\, X_1 = x)
\end{aligned} \tag{24}$$

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

$$\begin{aligned}
&\hat{\psi}^n_\eta(w',x') - \hat{\psi}^n_\eta(w'',x'') \\
&\quad = E\left(\hat{\phi}^{0:n}_{\eta,\boldsymbol{Y}}(w') - \hat{\phi}^{0:n}_{\eta,\boldsymbol{Y}}(e_0)\middle| X_1 = x'\right) \\
&\qquad - E\left(\hat{\phi}^{0:n}_{\eta,\boldsymbol{Y}}(w'') - \hat{\phi}^{0:n}_{\eta,\boldsymbol{Y}}(e_0)\middle| X_1 = x''\right) \\
&\qquad + \sum_{k=1}^{n-1}\sum_{x\in\mathcal{X}} E\left(\hat{\phi}^{k-1:n}_{\eta,\boldsymbol{Y}}(e_0) - \hat{\phi}^{k:n}_{\eta,\boldsymbol{Y}}(e_0)\middle| X_k = x\right) \\
&\qquad \cdot (p^{k-1}(x\,|\,x') - \pi(x)) \\
&\qquad - \sum_{k=1}^{n-1}\sum_{x\in\mathcal{X}} E\left(\hat{\phi}^{k-1:n}_{\eta,\boldsymbol{Y}}(e_0) - \hat{\phi}^{k:n}_{\eta,\boldsymbol{Y}}(e_0)\middle| X_k = x\right) \\
&\qquad \cdot (p^{k-1}(x\,|\,x'') - \pi(x)) \\
&\qquad + \sum_{x\in\mathcal{X}} E\left(\hat{\phi}^{n-1:n}_{\eta,\boldsymbol{Y}}(e_0)\,|\, X_n = x\right) \\
&\qquad \cdot (p^{n-1}(x\,|\,x') - \pi(x)) \\
&\qquad - \sum_{x\in\mathcal{X}} E(\hat{\phi}^{n-1:n}_{\eta,\boldsymbol{Y}}(e_0)\,|\, X_n = x) \\
&\qquad \cdot (p^{n-1}(x\,|\,x') - \pi(x))
\end{aligned} \tag{25}$$

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 $|\,p^n(x'\,|\,x) - \pi(x')\,| \leq \tilde{C}\tilde{\varepsilon}^n$ for each $x, x' \in \mathcal{X}$, $n \geq 0$.

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

$$\left|\hat{\phi}^{0:n}_{\eta,\boldsymbol{y}}(w)\right| \leq \psi_Q(y_n)$$

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

$$\left\|\nabla_\eta \hat{\phi}^{0:n}_{\eta,\boldsymbol{y}}(w)\right\| \leq \tilde{C}_{1,Q}\psi_Q(y_n) \tag{26}$$

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

$$E(\psi_Q(Y_n)\,|\,X_1 = x) \leq \max_{x'\in\mathcal{X}} \int \psi_Q(y')Q(dy'\,|\,x') < \infty \tag{27}$$

for all $x \in \mathcal{X}$, $n \geq 1$, it follows from the dominated convergence theorem and (26) that $\hat{\psi}^n_\eta(w,x)$ is differentiable (and thus, analytic) in η on $V_{\tilde{\delta}_{2,Q}}(Q)$ for any $w \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n \geq 1$.

Let $\tilde{\varepsilon}_Q = \max\{\varepsilon_{3,Q}, \tilde{\varepsilon}\}$. Due to Lemmas 1 and 4, we have

$$\begin{aligned}
&\left|\hat{\phi}^{0:n}_{\eta,\boldsymbol{y}}(w') - \hat{\phi}^{0:n}_{\eta,\boldsymbol{y}}(w'')\right| \\
&\quad \leq C_{1,Q}C_{8,Q}\varepsilon^{n-1}_{3,Q}\psi_Q(y_n)\|w' - w''\|
\end{aligned} \tag{28}$$

$$\begin{aligned}
&\left|\hat{\phi}^{k-1:n}_{\eta,\boldsymbol{y}}(w) - \hat{\phi}^{k:n}_{\eta,\boldsymbol{y}}(w)\right| \\
&\quad \leq C_{1,Q}\psi_Q(y_n)\left\|\hat{G}^{k:n-1}_{\eta,\boldsymbol{y}}(\hat{G}_\eta(w,y_k)) - \hat{G}^{k:n-1}_{\eta,\boldsymbol{y}}(w)\right\| \\
&\quad \leq C_{1,Q}C_{8,Q}\varepsilon^{n-k-1}_{3,Q}\psi_Q(y_n)\|\hat{G}_\eta(w,y_k) - w\|
\end{aligned} \tag{29}$$

for each $\eta \in V_{\tilde{\delta}_{2,Q}}(Q)$, $w, w', w'' \in V_{\tilde{\delta}_{2,Q}}(\mathcal{P}^{N_x})$, $n > 1$, $1 \leq k < n-1$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in $\mathcal{Y}$. Using (27)–(29), 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 the right-hand side of (25) is bounded by $\tilde{C}_{2,Q}\tilde{\varepsilon}^n_Q$ 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

$$|\,\hat{\psi}^n_\eta(w',x') - \hat{\psi}^n_\eta(w'',x'')\,| \leq 2\tilde{C}_{2,Q}\tilde{\varepsilon}^n_Q(n+1) \tag{30}$$

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, (24) yields

$$\begin{aligned}
&\left|\hat{\psi}^{n+1}_\eta(w,x) - \hat{\psi}^n_\eta(w,x)\right| \\
&\quad \leq E\left(\left|\hat{\psi}^n_\eta(\hat{G}_\eta(w,Y_1),X_2) - \hat{\psi}^n_\eta(w,x)\right|\middle| X_1 = x\right) \\
&\quad \leq 2\tilde{C}_{2,Q}\tilde{\varepsilon}^n_Q(n+1)
\end{aligned} \tag{31}$$

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 (30) and (31), there exists a function $\hat{\psi} : \mathbb{C}^{d_\theta} \to \mathbb{C}$ such that $\hat{\psi}^n_\eta(w,x)$ 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 [40, Th. 2.4.1]), $\hat{\psi}(\cdot)$ is analytic on $V_{\tilde{\delta}_{2,Q}}(Q)$. On the other side, since

$$\begin{aligned}\hat{\psi}^{n+1}_\theta(u,x) &= E\left(\phi_\theta(G^{0:n}_{\theta,\mathbf{Y}}(u), Y_{n+1}) \middle| X_1 = x\right)\\ &= E\left((\Pi^{n-1}\phi)(\theta,(x,Y_1,u)) \middle| X_1 = x\right)\end{aligned}$$

for all $\theta \in \Theta$, $u \in \mathcal{P}^{N_x}$, $x \in \mathcal{X}$, $n \geq 1$, Lemma 3 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., [20], [29], and [30]) and the analyticity of $f(\cdot)$. $\square$

As a direct consequence of [20, Th. ŁI, p. 775] and Theorem 1, we have the following corollary: $\square$

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

$$|f(\theta) - a| \leq M_{Q,a} \|\nabla f(\theta)\|^{\mu_{Q,a}} \tag{32}$$

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

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

D. Decomposition of Algorithm (3)

In this section, equivalent representations of recursion (3) are provided and their asymptotic properties are analyzed relying on the results of Section IV-B (Lemmas 1 and 3). The analysis is based on the techniques developed in [2, Part II]. The results of this section are a crucial prerequisite for the analysis carried out in Section IV-E. The following notation is used in this section. For $n \geq 0$, let $Z_{n+1} = (X_{n+1}, Y_{n+1}, U_n, V_n)$, while

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

and $\phi_n = \phi'_n + \phi''_n$ ($F(\cdot,\cdot)$ is defined in Section IV-B). Then, algorithm (3) admits the following representations for $n \geq 0$:

$$\begin{aligned}\theta_{n+1} &= \theta_n + \alpha_n F(\theta_n, Z_{n+1})\\ &= \theta_n + \alpha_n(\nabla f(\theta_n) + \xi_n).\end{aligned} \tag{33}$$

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 introduced in Section IV-B).

Lemma 5: 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

$$F(\theta, z) - \nabla f(\theta) = \Phi(\theta, z) - (P\Phi)(\theta, z) \tag{34}$$

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

$$\max\{\|F(\theta,z)\|, \|\Phi(\theta,z)\|, \|(P\Phi)(\theta,z)\|\} \leq \varphi_{Q,s}(z) \tag{35}$$

$$\|(P\Phi)(\theta',z) - (P\Phi)(\theta'',z)\| \leq \varphi_{Q,s}(z)\|\theta' - \theta''\|^s \tag{36}$$

$$\sup_{n\geq 0} E\left(\varphi^2_{Q,s}(Z_n) I_{\{\tau_Q \geq n\}} \middle| Z_0 = z\right) < \infty \tag{37}$$

for all $\theta, \theta', \theta'' \in Q$, $z \in \mathcal{S}_z$, where $\tau_Q = \inf(\{n \geq 0 : \theta_n \notin Q\} \cup \{\infty\})$.

Remark: Lemma 5 is a consequence of Lemmas 1–3 and the results of [2, ch. II.2]. It is proved in Appendix II.

Lemma 6: 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 $\Lambda \setminus N_0$.

Remark: The proof of Lemma 6 is based on the techniques developed in [2, ch. II.2]. It is provided in Appendix II.

Lemma 7: Suppose that Assumptions 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 (N_0 is specified in the statement of Lemma 6).

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

$$\begin{aligned}\varepsilon &\leq \|\nabla f(\theta_{n_k}) - \nabla f(\theta_{m_k})\| \\ &\leq \tilde{C}_Q \|\theta_{n_k} - \theta_{m_k}\| \\ &\leq \tilde{C}_Q^2 \sum_{i=m_k}^{n_k-1} \alpha_i + \tilde{C}_Q \left\| \sum_{i=m_k}^{n_k-1} \alpha_i \xi_i \right\| \end{aligned} \tag{38}$$

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 (38) would imply

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

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

E. Convergence and Convergence Rate

In this section, using the results of Sections IV-C and IV-D (Corollary 5 and Lemmas 5 and 6), the convergence and convergence rate of recursion (3) are analyzed and Theorems 2 and 3 are proved.

Throughout this section, 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

$$\begin{aligned}\zeta_n &= \sup_{k\geq n} \left\| \sum_{i=n}^{k} \alpha_i \xi_i \right\| \\ \varepsilon'_{n,k} &= \sum_{i=n}^{k-1} \alpha_i \xi_i \\ \varepsilon''_{n,k} &= \sum_{i=n}^{k-1} \alpha_i (\nabla f(\theta_i) - \nabla f(\theta_n)).\end{aligned}$$

For the same n, k, let

$$\begin{aligned}\phi'_{n,k} &= (\nabla f(\theta_n))^T (\varepsilon'_{n,k} + \varepsilon''_{n,k}) \\ \phi''_{n,k} &= \int_0^1 (\nabla f(\theta_n + t(\theta_k - \theta_n)) - \nabla f(\theta_n))^T (\theta_k - \theta_n) dt\end{aligned}$$

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

$$\begin{aligned}\theta_k &= \theta_n + \sum_{i=n}^{k-1} \alpha_i \nabla f(\theta_i) + \varepsilon'_{n,k} \\ &= \theta_n + (\gamma_k - \gamma_n) \nabla f(\theta_n) + \varepsilon_{n,k} \end{aligned} \tag{39}$$

$$f(\theta_k) = f(\theta_n) + (\gamma_k - \gamma_n) \|\nabla f(\theta_n)\|^2 + \phi_{n,k} \tag{40}$$

for $0 \leq n \leq k$.

Besides the notation introduced in the previous paragraph, we also rely on the following notation in this section. 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 the random quantity and the random sets (respectively) defined by

$$\begin{aligned}\hat{\rho} &= d(\hat{A}, \partial\Theta)/2 \\ \hat{B} &= \bigcup_{\theta\in\hat{A}} \{\theta' \in \mathbb{R}^{d_\theta} : \|\theta' - \theta\| \leq \min\{\delta_\theta, \hat{\rho}\}\} \\ \hat{Q} &= \mathrm{cl}(\hat{B})\end{aligned}$$

on Λ, and by

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

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

$$\hat{\delta} = \delta_{\hat{Q},\hat{f}}, \quad \hat{\mu} = \mu_{\hat{Q},\hat{f}}, \quad \hat{C} = C_{\hat{Q}}, \quad \hat{M} = M_{\hat{Q},\hat{f}} \tag{41}$$

on Λ, and as

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

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 (41) are equivalent]. Random quantities $\hat{p}$, $\hat{q}$, $\hat{r}$ are defined in the same way as in (9)–(11). Functions $u(\cdot)$ and $v(\cdot)$ are defined by

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

for $\theta \in \Theta$.

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

$$|f(\theta) - \hat{f}| \leq \hat{M} \|\nabla f(\theta)\|^{\hat{\mu}} \tag{42}$$

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

Remark: Throughout this section, the following convention is applied. Diacritic ˜ is used to denote a locally defined quantity, i.e., a quantity whose definition holds only in the proof where the quantity appears.

Lemma 8: 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 6).

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

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

for $0\leq n\leq k$. Consequently, Lemma 6 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 9: 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

$$\max_{n\leq k\leq a(n,\hat{t})}\|\varepsilon_{n,k}\|\leq(\hat{t}/\hat{C}_1)(\gamma_n^{-r}+\|\nabla f(\theta_n)\|)\quad(43)$$

$$\max_{n\leq k\leq a(n,\hat{t})}|\,\phi_{n,k}\,|\leq(\hat{t}/\hat{C}_1)(\gamma_n^{-2r}+\|\nabla f(\theta_n)\|^2)\quad(44)$$

$$f(\theta_n)-f(\theta_{a(n,\hat{t})})+2^{-1}\hat{t}\|\nabla f(\theta_n)\|^2\leq(\hat{t}/\hat{C}_1)\gamma_n^{-2r}\quad(45)$$

$$\begin{aligned}&f(\theta_n)-f(\theta_{a(n,\hat{t})})+2^{-1}\|\nabla f(\theta_n)\|\|\theta_{a(n,\hat{t})}-\theta_n\|\\&\quad\leq(\hat{t}/\hat{C}_1)\gamma_n^{-2r}\end{aligned}\quad(46)$$

on $\Lambda\setminus N_0$ for $n>\sigma$ (N_0 is specified in the statement of Lemma 6).

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

$$\begin{aligned}\tilde{\sigma}_1&=\max(\{n\geq0:\theta_n\notin\hat{Q}\}\cup\{0\})\\\tilde{\sigma}_2&=\max(\{n\geq0:\alpha_n>\hat{t}/4\}\cup\{0\})\\\tilde{\sigma}_3&=\max(\{n\geq0:\gamma_n^r\zeta_n>\hat{t}/(2\hat{C}_1\tilde{C}_4)\}\cup\{0\})\end{aligned}$$

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 8 implies $0\leq\sigma<\infty$ everywhere. We also have

$$\begin{aligned}&\max\left\{\tilde{C}_2\gamma_n^r\zeta_n,\tilde{C}_3\gamma_n^r\zeta_n,\tilde{C}_3\gamma_n^{2r}\zeta_n^2,\tilde{C}_4\gamma_n^r\zeta_n,\tilde{C}_4\gamma_n^{2r}\zeta_n^2\right\}\\&\quad\leq2^{-1}\hat{C}_1^{-1}\hat{t}\end{aligned}\quad(47)$$

$$\max\{\tilde{C}_2\hat{t}^2,\tilde{C}_3\hat{t}^2,\tilde{C}_4\hat{t}^2\}\leq2^{-1}\hat{C}_1^{-1}\hat{t}\quad(48)$$

$$\hat{t}\geq\gamma_{a(n,\hat{t})}-\gamma_n=\gamma_{a(n,\hat{t})+1}-\gamma_n-\alpha_{a(n,\hat{t})}\geq3\hat{t}/4\quad(49)$$

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

Let ω be an arbitrary sample from $\Lambda\setminus N_0$ (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

$$\begin{aligned}\|\nabla f(\theta_k)\|&\leq\hat{C}(\zeta_n+\|\nabla f(\theta_n)\|)\exp(\hat{C}(\gamma_k-\gamma_n))\\&\leq\hat{C}\exp(\hat{C})(\zeta_n+\|\nabla f(\theta_n)\|)\end{aligned}\quad(50)$$

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

$$\begin{aligned}\|\varepsilon_{n,k}\|&\leq\|\varepsilon'_{n,k}\|+\hat{C}\sum_{i=n}^{k-1}\alpha_i\|\theta_i-\theta_n\|\\&\leq\zeta_n+\hat{C}\tilde{C}_1((\gamma_k-\gamma_n)\zeta_n+(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|)\\&\leq\tilde{C}_2(\zeta_n+(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|)\end{aligned}\quad(51)$$

for $\sigma<n\leq k\leq a(n,1)$ (notice that $\gamma_k-\gamma_n\leq1$ for $n\leq k\leq a(n,1)$). Thus

$$\begin{aligned}|\,\phi_{n,k}\,|&\leq\|\nabla f(\theta_n)\|\|\varepsilon_{n,k}\|+\hat{C}\|\theta_k-\theta_n\|^2\\&\leq\tilde{C}_2(\zeta_n\|\nabla f(\theta_n)\|+(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|^2)\\&\quad+2\hat{C}\tilde{C}_1^2(\zeta_n^2+(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|^2)\\&\leq\tilde{C}_3(\zeta_n^2+\zeta_n\|\nabla f(\theta_n)\|+(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|^2)\end{aligned}\quad(52)$$

for $\sigma<n\leq k\leq a(n,1)$. On the other side, combining (39) and (40), we get

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

for $0\leq n\leq k$. Then, (51) and (52) yield

$$\begin{aligned}&f(\theta_n)-f(\theta_k)+\|\nabla f(\theta_n)\|\|\theta_k-\theta_n\|\\&\quad\leq\|\nabla f(\theta_n)\|\|\varepsilon_{n,k}\|+|\,\phi_{n,k}\,|\\&\quad\leq\tilde{C}_3\zeta_n^2+(\tilde{C}_2+\tilde{C}_3)\zeta_n\|\nabla f(\theta_n)\|\\&\qquad+(\tilde{C}_2+\tilde{C}_3)(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|^2\\&\quad\leq\tilde{C}_4(\zeta_n^2+\zeta_n\|\nabla f(\theta_n)\|+(\gamma_k-\gamma_n)^2\|\nabla f(\theta_n)\|^2)\end{aligned}\quad(53)$$

for $\sigma < n \leq k \leq a(n,1)$. Owing to (47), (48), (51), and (52), we have

$$\begin{aligned}\|\varepsilon_{n,k}\| &\leq \tilde{C}_2\zeta_n + \tilde{C}_2\hat{t}^2\|\nabla f(\theta_n)\| \\ &\leq \hat{C}_1^{-1}\hat{t}(\gamma_n^{-r} + \|\nabla f(\theta_n)\|) \end{aligned} \tag{54}$$

$$\begin{aligned}|\,\phi_{n,k}\,| &\leq \tilde{C}_3\zeta_n^2 + \tilde{C}_3\zeta_n\|\nabla f(\theta_n)\| + \tilde{C}_3\hat{t}^2\|\nabla f(\theta_n)\|^2 \\ &\leq 2^{-1}\hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \gamma_n^{-r}\|\nabla f(\theta_n)\| + \|\nabla f(\theta_n)\|^2) \\ &\leq \hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2) \end{aligned} \tag{55}$$

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 (40), (49), and (55), we have also

$$\begin{aligned} &f\left(\theta_n\right) - f(\theta_{a(n,\hat{t})}) \\ &\quad\leq -(\gamma_{a(n,\hat{t})} - \gamma_n)\|\nabla f(\theta_n)\|^2 + |\,\phi_{n,a(n,\hat{t})}\,| \\ &\quad\leq -(3\hat{t}/4)\|\nabla f(\theta_n)\|^2 + \hat{C}_1^{-1}\hat{t}(\gamma_n^{-2r} + \|\nabla f(\theta_n)\|^2) \\ &\quad= -(3/4 - \hat{C}_1^{-1})\hat{t}\|\nabla f(\theta_n)\|^2 + \hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} \\ &\quad\leq -2^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 + \hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} \end{aligned} \tag{56}$$

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

$$\begin{aligned}\hat{C}_1^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 &\leq 2^{-1}\hat{t}\|\nabla f(\theta_n)\|^2 \\ &\leq \hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} + (f(\theta_{a(n,\hat{t})}) - f(\theta_n)) \end{aligned} \tag{57}$$

for $n > \sigma$. On the other side, (47)–(49), (53), and (57) imply

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

for $n > \sigma$. Therefore

$$2(f(\theta_n) - f(\theta_{a(n,\hat{t})})) + \|\nabla f(\theta_n)\|\|\theta_{a(n,\hat{t})} - \theta_n\| \leq 2\hat{C}_1^{-1}\hat{t}\gamma_n^{-2r} \tag{58}$$

for $n > \sigma$. Then, (43)–(46) directly follow from (54), (55), (56), (57), and (58). □

Lemma 10: 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

$$\left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/4)\|\nabla f(\theta_n)\|^2\right) I_{A_n} \leq 0 \tag{59}$$

$$\left(u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/\hat{C}_2)u(\theta_n)\right) I_{B_n} \leq 0 \tag{60}$$

$$\left(v(\theta_{a(n,\hat{t})}) - v(\theta_n) - \hat{t}/\hat{C}_2\right) I_{C_n} \geq 0 \tag{61}$$

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

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

(N_0 and $\hat{t}$ are specified in the statements of Lemmas 6 and 9, respectively).

Remark: Inequalities (59)–(61) can be interpreted in the following way: relations

$$\begin{aligned} &(\gamma_n^{\hat{p}}\,|\,u(\theta_n)\,| \geq 1 \vee \gamma_n^r\|\nabla f(\theta_n)\| \geq 1) \wedge n > \tau \\ &\qquad\Longrightarrow u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/4)\|\nabla f(\theta_n)\|^2 \end{aligned} \tag{62}$$

$$\begin{aligned} &\gamma_n^{\hat{p}}u(\theta_n)\,| \geq 1 \wedge \hat{\mu} = 2 \wedge n > \tau \\ &\qquad\Longrightarrow u(\theta_{a(n,\hat{t})}) \leq (1 - \hat{t}/\hat{C}_2)u(\theta_n) \end{aligned} \tag{63}$$

$$\begin{aligned} &\gamma_n^{\hat{p}}u(\theta_n) \geq 1 \wedge \hat{\mu} < 2 \wedge n > \tau \\ &\qquad\Longrightarrow v(\theta_{a(n,\hat{t})}) - v(\theta_n) \geq \hat{t}/\hat{C}_2 \end{aligned} \tag{64}$$

are true on $\Lambda \setminus N_0$.

Proof: Let

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

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 7 implies $0 \leq \tau < \infty$ everywhere. On the other side, since $\tau \geq \sigma$ on $\Lambda \setminus N_0$, Lemma 9 [inequality (45)] implies

$$u(\theta_{a(n,\hat{t})}) - u(\theta_n) \leq -(\hat{t}/2)\|\nabla f(\theta_n)\|^2 + (\hat{t}/\hat{C}_1)\gamma_n^{-2r} \tag{65}$$

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 $n > \tau$, (42) (i.e., Corollary 5) yields

$$|\,u(\theta_n)\,| \leq \hat{M}\|\nabla f(\theta_n)\|^{\hat{\mu}} \tag{66}$$

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 (59). We proceed by contradiction: suppose that (59) is violated for some $n > \tau$. Consequently

$$u(\theta_{a(n,\hat{t})}) - u(\theta_n) + (\hat{t}/4)\|\nabla f(\theta_n)\|^2 > 0 \tag{67}$$

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

$$|\,u(\theta_n)\,| \geq \gamma_n^{-\hat{p}}, \quad \|\nabla f(\theta_n)\| \geq \gamma_n^{-r}. \tag{68}$$

If $|\,u(\theta_n)\,| \geq \gamma_n^{-\hat{p}}$, then (66) implies

$$\begin{aligned}\|\nabla f(\theta_n)\|^2 &\geq (\,|\,u(\theta_n)\,|\,/\hat{M})^{2/\hat{\mu}} \\ &\geq (1/\hat{M})^{2/\hat{\mu}}\gamma_n^{-2\hat{p}/\hat{\mu}} \\ &\geq (4/\hat{C}_1)\gamma_n^{-2r}\end{aligned}$$

(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 two inequalities in (68), 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, (65) implies

$$u\left(\theta_{a(n,\hat{t})}\right) - u(\theta_n) \leq -(\hat{t}/4)\|\nabla f(\theta_n)\|^2 \tag{69}$$

which directly contradicts (67). Hence, (59) is true for $n > \tau$. Owing to this, (66), and the fact that $B_n \subset A_n$ for $n \geq 0$, we obtain

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

for $n > \tau$ (notice that $u(\theta_n) > 0$ on B_n; also notice that $4\hat{M} \leq \hat{C}_2$). Thus, (60) is satisfied.

Now, let us prove (61). To do so, we again use contradiction: suppose that (61) does not hold for some $n > \tau$. Consequently, we have $\hat{\mu} < 2$, $u(\theta_{a(n,\hat{t})}) > 0$ and

$$\gamma_n^{\hat{p}} u(\theta_n) \geq 1 \tag{70}$$

$$v\left(\theta_{a(n,\hat{t})}\right) - v(\theta_n) < \hat{t}/\hat{C}_2. \tag{71}$$

Combining (70) with (already proved) (59), we get (69). On the other side, (66) yields

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

(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, (69) 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 (71). Hence, (60) is satisfied for $n > \tau$. ☐

Lemma 11: Suppose that Assumptions 1–4 hold. Then

$$\gamma_n^{\hat{p}} u(\theta_n) \geq -1 \tag{72}$$

$$\|\nabla f(\theta_n)\|^2 \leq (4/\hat{t})\left(\varphi(u(\theta_n)) + \gamma_n^{-\hat{p}}\right) \tag{73}$$

on $\Lambda \setminus N_0$ for $n > \tau$, where function $\varphi(\cdot)$ is defined by $\varphi(x) = x\mathrm{I}_{(0,\infty)}(x)$, $x \in \mathbb{R}$ (N_0 is specified in the statement of Lemma 6).

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 (72). To do so, we use contradiction: assume that (72) 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 nonincreasing: 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 10 [relations (59) and (62)] yields

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

i.e., $u(\theta_{n_{k+1}}) \leq u(\theta_{n_k})$. Thus, $\{u(\theta_{n_k})\}_{k\geq 0}$ is nonincreasing. 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 7). Hence, (72) indeed holds for $n > \tau$.

Now, (73) is demonstrated. Again, we proceed by contradiction: suppose that (73) 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 10 [relations (59) and (62)], yields

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

Then, (72) implies

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

However, this directly contradicts our assumption that n violates (73). Thus, (73) is satisfied for $n > \tau$. ☐

Lemma 12: Suppose that Assumptions 1–4 hold. Let $\hat{C}_3 = 2\hat{C}_2^{\hat{p}}$. Then

$$\liminf_{n\to\infty} \gamma_n^{\hat{p}} u(\theta_n) \leq \hat{C}_3 \tag{74}$$

on $\Lambda \setminus N_0$ (N_0 is specified in the statement of Lemma 6).

Proof: We prove the lemma by contradiction: assume that (74) 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

$$\gamma_n^{\hat{p}} u(\theta_n) \geq \hat{C}_3 \tag{75}$$

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 10 [relations (61) and (64)] and (75), we have

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

for $k \geq 0$ [notice that $\gamma_n^{\hat{p}} u(\theta_n) \geq 1$ due to (75); also notice that $\gamma_{n_{k+1}} - \gamma_{n_k} \leq \hat{t}$]. Therefore

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

for $k \geq 0$. Then, (75) implies

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

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, (74) holds when $\hat{\mu} < 2$.

Case $\hat{\mu} = 2$: Due to Lemma 10 [relations (60) and (63)] and (75), we have

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

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, (75) 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, (74) holds in the case $\hat{\mu} = 2$ as well. □

Lemma 13: Suppose that Assumptions 1–4 hold. Let $\hat{C}_4 = 6\hat{C}_3$. Then

$$\limsup_{n \to \infty} \gamma_n^{\hat{p}} u(\theta_n) \leq \hat{C}_4 \tag{76}$$

on $\Lambda \setminus N_0$ (N_0 is specified in the statement of Lemma 6).

Proof: We use contradiction to prove the lemma: suppose that (76) 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 12 that there exist $n_0 > m_0 > \tau$ such that

$$\gamma_{m_0}^{\hat{p}} u(\theta_{m_0}) \leq 2\hat{C}_3 \tag{77}$$

$$\gamma_{n_0}^{\hat{p}} u(\theta_{n_0}) > \hat{C}_4 \tag{78}$$

$$\min_{m_0 < n \leq n_0} \gamma_n^{\hat{p}} u(\theta_n) > 2\hat{C}_3 \tag{79}$$

$$\max_{m_0 \leq n < n_0} \gamma_n^{\hat{p}} u(\theta_n) \leq \hat{C}_4 \tag{80}$$

and such that

$$(\gamma_{a(m_0,\hat{t})}/\gamma_{m_0})^{\hat{p}} \leq \min\{2, (1 - \hat{t}/\hat{C}_2)^{-1}\}. \tag{81}$$

Let $l_0 = a(m_0, \hat{t})$. As a direct consequence of Lemma 11 and (77), 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 9 and (40) 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, (77) and (79) yield

$$\begin{aligned} u(\theta_{m_0}) &\geq u(\theta_{m_0+1}) - \gamma_{m_0}^{-\hat{p}} \\ &\geq 2\hat{C}_3 (\gamma_{m_0}/\gamma_{m_0+1})^{\hat{p}} \gamma_{m_0}^{-\hat{p}} - \gamma_{m_0}^{-\hat{p}} \\ &\geq (\hat{C}_3 - 1)\gamma_{m_0}^{-\hat{p}} \geq \gamma_{m_0}^{-\hat{p}} \end{aligned} \tag{82}$$

$$\begin{aligned} u(\theta_n) &\leq u(\theta_{m_0}) + \gamma_{m_0}^{-\hat{p}} \\ &\leq (2\hat{C}_3 + 1)(\gamma_n/\gamma_{m_0})^{\hat{p}} \gamma_n^{-\hat{p}} \\ &\leq 6\hat{C}_3 \gamma_n^{-\hat{p}} = \hat{C}_4 \gamma_n^{-\hat{p}} \end{aligned} \tag{83}$$

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 (78) and (83), 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 10 [relations (61) and (64)] and (77), (82), 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 (79) and the fact that $m_0 < l_0 < n_0$. Thus, (76) holds when $\hat{\mu} < 2$.

Case $\hat{\mu} = 2$: Using Lemma 10 [relations (60) and (63)] and (82), we get

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

However, this is impossible due to (79) and the fact that $m_0 < l_0 < n_0$. Hence, (76) holds in the case $\hat{\mu} = 2$ as well. □

Lemma 14: Suppose that Assumptions 1–4 hold. Then

$$\|\theta_{a(n,\hat{t})} - \theta_n\| \leq 2\gamma_n^{\hat{q}+1}(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + 6\gamma_n^{-(\hat{q}+1)} \tag{84}$$

on $\Lambda \setminus N_0$ for $n > \max\{\sigma, \tau\}$ (N_0 is specified in the statement of Lemma 6).

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 (84), 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 9, we have

$$\|\nabla f(\theta_n)\| \|\theta_{a(n,\hat{t})} - \theta_n\| \leq 2(u(\theta_n) - u(\theta_{a(n,\hat{t})})) + 2(\hat{t}/\hat{C}_1)\gamma_n^{-2r}. \tag{85}$$

On the other side, since $\|\nabla f(\theta_n)\| \geq \gamma_n^{-(\hat{q}+1)} \geq \gamma_n^{-r}$ (notice that $\hat{q} + 1 \leq r$), Lemma 10 [relations (59) and (62)] 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, (85) yields

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

[notice that $\hat{t}/\hat{C}_1 \le 1/2$; also notice that $\hat{q}+1 \le r$, which implies $2r - (\hat{q}+1) \ge \hat{q}+1$]. Hence, (84) is true when $\|\nabla f(\theta_n)\| \ge \gamma_n^{-(\hat{q}+1)}$.

Case $\|\nabla f(\theta_n)\| < \gamma_n^{-(\hat{q}+1)}$: Using Lemma 9 and (40), we get

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

(notice that $\hat{q}+1 \le r < 2r$ and $\hat{t}/\hat{C}_1 \le 1/2$). On the other side, owing to Lemma 9 and (39), we have

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

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

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

Thus, (84) holds in the case $\|\nabla f(\theta_n)\| < \gamma_n^{-(\hat{q}+1)}$. □

Lemma 15: Suppose that Assumptions 1–4 hold. Then, there exists a random quantity $\hat{C}_5$ such that $1 \le \hat{C}_5 < \infty$ everywhere and such that

$$\limsup_{n\to\infty} \gamma_n^{\hat{q}} \max_{k\ge n} \|\theta_k - \theta_n\| \le \hat{C}_5 \tag{86}$$

on $\Lambda \setminus N_0$ (N_0 is specified in the statement of Lemma 6).

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).

First, we show

$$\hat{q} = \min\{(\hat{p}+1)/2, r\} - 1. \tag{87}$$

If $\hat{\mu} = 2$, then $\hat{r} = \infty$, $\hat{p} = 2r$, $\hat{q} = r-1$, and consequently, (87) holds. If $\hat{\mu} < 2$, $r > \hat{r}$, then $\hat{r} = 1/(2-\hat{\mu})$, $\hat{p} = \hat{\mu}\hat{r}$, $\hat{q} = \hat{r} - 1$ and thus, $(\hat{p}+1)/2 = \hat{r}$. Therefore, (87) is true when $\hat{\mu} < 2$, $r > \hat{r}$. If $\hat{\mu} < 2$, $r \le \hat{r}$, then $\hat{r} = 1/(2-\hat{\mu})$, $\hat{p} = \hat{\mu}r$, $\hat{q} = r-1$, and hence, $\hat{p}+1 = \hat{\mu}r + 1 \ge 2r$ (notice that $r \le \hat{r}$, $\hat{r} = 1/(2-\hat{\mu})$ imply $2r - \hat{\mu}r \le 1$). Consequently, (87) is true when $\hat{\mu} < 2$, $r \le \hat{r}$.

Using Lemmas 11 and 13, we get

$$\limsup_{n\to\infty} \gamma_n^{\hat{p}}\,|\,u(\theta_n)\,| \le \hat{C}_4 \tag{88}$$

$$\limsup_{n\to\infty} \gamma_n^{\hat{p}}\|\nabla f(\theta_n)\|^2 \le 8\hat{C}_4/\hat{t}. \tag{89}$$

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 (88) and (89) that there exists $n_0 > \max\{\sigma, \tau\}$ (depending on ω) such that $|\,u(\theta_n)\,| \le 2\hat{C}_4\gamma_n^{-\hat{p}}$, $\|\nabla f(\theta_n)\| \le (4\hat{C}_4/\hat{t})\gamma_n^{-\hat{p}/2}$, $\gamma_{a(n,\hat{t})} - \gamma_n \ge \hat{t}/2$, and

$$(1 - \hat{t}/\gamma_n)^{\hat{q}+1} \ge 1 - (\hat{q}+1)\gamma_n^{-1} \tag{90}$$

for $n \ge n_0$. Then, (39) and Lemma 9 imply

$$\begin{aligned}\|\theta_k - \theta_n\| &\le (\gamma_k - \gamma_n)\|\nabla f(\theta_n)\| + \|\varepsilon_{n,k}\| \\ &\le \hat{t}\|\nabla f(\theta_n)\| + (\hat{t}/\hat{C}_1)(\gamma_n^{-r} + \|\nabla f(\theta_n)\|) \\ &\le 8\hat{C}_4\gamma_n^{-\hat{p}/2} + \gamma_n^{-r} \\ &\le \tilde{C}\gamma_n^{-\hat{q}}\end{aligned} \tag{91}$$

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

Let $\{n_k\}_{k\ge 0}$ be recursively defined as $n_{k+1} = a(n_k, \hat{t})$ for $k \ge 0$. Due to Lemma 14, we have

$$\begin{aligned}&\|\theta_{n_l} - \theta_{n_k}\| \\ &\quad\le \sum_{i=k}^{l-1} \|\theta_{n_{i+1}} - \theta_{n_i}\| \\ &\quad\le 2\sum_{i=k}^{l-1} \gamma_{n_i}^{\hat{q}+1}(u(\theta_{n_i}) - u(\theta_{n_{i+1}})) + 6\sum_{i=k}^{l-1}\gamma_{n_i}^{-(\hat{q}+1)} \\ &\quad\le 2\sum_{i=k+1}^{l} (\gamma_{n_i}^{\hat{q}+1} - \gamma_{n_{i-1}}^{\hat{q}+1})\,|\,u(\theta_{n_i})\,| \\ &\qquad + 2\gamma_{n_l}^{\hat{q}+1}\,|\,u(\theta_{n_l})\,| + 2\gamma_{n_k}^{\hat{q}+1}\,|\,u(\theta_{n_k})\,| + 6\sum_{i=k}^{l-1}\gamma_{n_i}^{-(\hat{q}+1)}\end{aligned} \tag{92}$$

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

$$\begin{aligned}\gamma_{n_i}^{\hat{q}+1} - \gamma_{n_{i-1}}^{\hat{q}+1} &= \gamma_{n_i}^{\hat{q}+1}(1 - (1 - (\gamma_{n_i} - \gamma_{n_{i-1}})/\gamma_{n_i})^{\hat{q}+1}) \\ &\le \gamma_{n_i}^{\hat{q}+1}(1 - (1 - \hat{t}/\gamma_{n_i})^{\hat{q}+1}) \\ &\le (\hat{q}+1)\gamma_{n_i}^{\hat{q}}\end{aligned}$$

for $i \ge 0$ [use (90)], we get

$$\sum_{i=k+1}^{l} (\gamma_{n_i}^{\hat{q}+1} - \gamma_{n_{i-1}}^{\hat{q}+1})\gamma_{n_i}^{-\hat{p}} \le (\hat{q}+1)\sum_{i=k}^{\infty}\gamma_{n_i}^{-\hat{p}+\hat{q}} \tag{93}$$

$$\le (\hat{q}+1)\sum_{i=k}^{\infty}\gamma_{n_i}^{-(\hat{q}+1)} \tag{94}$$

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

$$\begin{aligned}\sum_{i=k}^{\infty} \gamma_{n_i}^{-(\hat{q}+1)} &\leq \sum_{i=0}^{\infty} (\gamma_{n_k} + \hat{t}i/2)^{-(\hat{q}+1)} \\ &\leq \gamma_{n_k}^{-(\hat{q}+1)} + \int_0^{\infty} (\gamma_{n_k} + \hat{t}u/2)^{-(\hat{q}+1)} du \\ &= \gamma_{n_k}^{-(\hat{q}+1)} + 2\hat{t}^{-1}\hat{q}^{-1}\gamma_{n_k}^{-\hat{q}} \\ &\leq (1 + 2\hat{t}^{-1}\hat{q}^{-1})\gamma_{n_k}^{-\hat{q}} \end{aligned} \tag{95}$$

for $k \geq 0$. Consequently, (92) and (93) imply

$$\begin{aligned}\|\theta_{n_l} - \theta_{n_k}\| &\leq (6 + 4\hat{C}_4(\hat{q}+1)) \sum_{i=k}^{\infty} \gamma_{n_i}^{-(\hat{q}+1)} \\ &\quad + 4\hat{C}_4\gamma_{n_k}^{-\hat{p}+\hat{q}+1} + 4\hat{C}_4\gamma_{n_l}^{-\hat{p}+\hat{q}+1} \\ &\leq 16\tilde{C}(1 + \hat{t}^{-1}\hat{q}^{-1})\gamma_{n_k}^{-\hat{q}} \end{aligned} \tag{96}$$

for $l \geq k \geq 0$ [notice that $\hat{p} - (\hat{q}+1) \geq (\hat{p}-1)/2 \geq \hat{q}$]. Using (91) and (96), 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 (86) is true. □

Proof of Theorems 2 and 3: Owing to Lemmas 7 and 15, 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 section coincide with $\hat{p}$, $\hat{q}$ introduced in Theorem 3 (see the remark after Corollary 5). Then, Lemmas 11, 13, and 15 imply that (6)–(8) are true on $\Lambda \setminus N_0$. □

V. Proof of Propositions 1–4

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

$$\varepsilon_Q \leq r_\theta(y, x' \,|\, x) \leq \varepsilon_Q^{-1} \tag{97}$$

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 a (complex-valued) analytic 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 analytic 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

$$\hat{\phi}_\eta(w, y) = \begin{cases} \log(e^T \hat{R}_\eta(y) w), & \text{if } e^T \hat{R}_\eta(y) w \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{98}$$

$$\hat{G}_\eta(w, y) = \begin{cases} \hat{R}_\eta(y) w / (e^T \hat{R}_\eta(y) w), & \text{if } e^T \hat{R}_\eta(y) w \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{99}$$

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 (97)], 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 analytic 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 as well. □

Proof of Proposition 2: Conditions i) and 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 a (complex-valued) analytic 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 analytic 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 (98) and (99) 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 analytic 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 analytic 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 as well. □

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 a (complex-valued) analytic 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 analytic 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 (98) and (99). 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 $\operatorname{Re}\{\hat{h}_{\eta,w}^k(x')\} \geq \varepsilon_Q/2$, $|\,\hat{h}_{\eta,w}^k(x')\,| \leq 2\varepsilon_Q^{-1}$ 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

$$\begin{aligned}
&|\,e^T \hat{R}_\eta(y) w\,| \\
&\quad\geq |\operatorname{Re}\{e^T \hat{R}_\eta(y) w\}| \\
&\quad= \sum_{x' \in \mathcal{X}} \sum_{k=1}^{N_\beta} f_k(y\,|\,x') \operatorname{Re}\{\hat{h}_{\eta,w}^k(x')\} \\
&\quad\geq (\varepsilon_Q/2)\psi(y) > 0 \\
&\max\{\|\hat{R}_\eta(y) w\|, |\,e^T \hat{R}_\eta(y) w\,|\} \\
&\quad\leq \sum_{x' \in \mathcal{X}} \sum_{k=1}^{N_\beta} f_k(y\,|\,x') \,|\,\hat{h}_{\eta,w}^k(x')\,| \\
&\quad\leq 2\varepsilon_Q^{-1}\psi(y)
\end{aligned}$$

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 analytic in (η, w) on $V_{\delta_Q}(Q) \times V_{\delta_Q}(\mathcal{P}^{N_x})$ for each $y \in \mathcal{Y}$. Moreover

$$\begin{aligned}
\|\hat{G}_\eta(w,y)\| &\leq 4\varepsilon_Q^{-2} \\
|\,\hat{\phi}_\eta(w,y)\,| &\leq |\,\log|\,e^T \hat{R}_\eta(y) w\,|\,| + 2\pi \\
&\leq |\,\log \psi(y)\,| + \log(2\varepsilon_Q^{-1}) + 2\pi
\end{aligned}$$

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 as well. □

Lemma 16: Let the conditions of Proposition 4 hold. Then, $\phi_\theta(u,y)$, $G_\theta(u,y)$ have (complex-valued) analytic 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 analytic 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

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

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 a (complex-valued) analytic 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 analytic in a on $V_{\tilde{\delta}_\alpha}(\alpha)$ for each $x, x' \in \mathcal{X}$.

On the other side, the analytic 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 (98) and (99).

Let α, $\beta = [\lambda_1 \cdots \lambda_{N_x} \mu_1 \cdots \mu_{N_x}]^T$ be arbitrarily vectors in $\mathcal{A}$, $\mathcal{B}$ (respectively), while $\theta = [\alpha^T \beta^T]^T$. Obviously, it can be assumed without loss of generality that $\lambda_1 < \lambda_x$ for each $x \in \mathcal{X} \setminus \{1\}$. 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 analytic in η on $V_{\tilde{\delta}_{1,\theta}}(\theta)$ for any $y \in \mathcal{Y}$, and such that

$$\operatorname{Re}\left\{\sum_{x \in \mathcal{X}} \hat{p}_a(x'\,|\,x) w_x\right\} \geq \tilde{\varepsilon}_\theta \tag{100}$$

$$\left|\sum_{x \in \mathcal{X}} \hat{p}_a(x'\,|\,x) w_x\right| \leq \tilde{\varepsilon}_\theta^{-1} \tag{101}$$

$$\min\{\operatorname{Re}\{l_1\}, \operatorname{Re}\{l_{x'} - l_1\}\} \geq \tilde{\varepsilon}_\theta$$
$$\max\{|\,l_{x''}\,|, |\,m_{x''}\,|\} \leq \tilde{\varepsilon}_\theta^{-1}$$

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

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

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

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

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

$$\left|\frac{\hat{q}_b(y\,|\,x)}{\hat{q}_b(y\,|\,1)}\right| \le \tilde{C}_\theta, \quad |\,\log|\,\hat{q}_b(y\,|\,x)\,|\,| \le \tilde{C}_\theta(1+y^2) \tag{102}$$

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

$$\left|\frac{\hat{q}_b(y\,|\,x)}{\hat{q}_b(y\,|\,1)}\right| \le 2^{-1}N_x^{-1}\tilde{\varepsilon}_\theta^2 \tag{103}$$

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 (103) 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

$$\left|\frac{\hat{q}_b(y\,|\,x)}{q_\beta(y\,|\,x)} - 1\right| \le 2^{-1}\tilde{\varepsilon}_\theta^2, \quad \left|\frac{\hat{q}_b(y\,|\,x)}{q_\beta(y\,|\,x)}\right| \le 2 \tag{104}$$

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

$$|\hat{q}_b(y\,|\,x) - q_\beta(y\,|\,x)| \le 2^{-1}\tilde{\varepsilon}_\theta^2 q_\beta(y\,|\,x) \tag{105}$$

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

Let $\delta_\theta = \min\{\tilde{\delta}_{1,\theta}, \tilde{\delta}_{2,\theta}\}$, $K_\theta = 8N_x\tilde{C}_\theta\tilde{\varepsilon}_\theta^{-2}$. As a result of (101) and (102), we have

$$\begin{aligned}
&\max\{\|\hat{R}_\eta(y)w\|, |\,e^T\hat{R}_\eta(y)w\,|\} \\
&\quad\le \sum_{x'\in\mathcal{X}} |\,\hat{q}_b(y\,|\,x')\,| \left|\sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right| \\
&\quad\le N_x\tilde{C}_\theta\tilde{\varepsilon}_\theta^{-1}|\,\hat{q}_b(y\,|\,1)\,|
\end{aligned} \tag{106}$$

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 (100), (101), and (103), we get

$$\begin{aligned}
&\frac{|\,e^T\hat{R}_\eta(y)w\,|}{|\,\hat{q}_b(y\,|\,1)\,|} \\
&\quad= \left|\sum_{x'\in\mathcal{X}} \frac{\hat{q}_b(y\,|\,x')}{\hat{q}_b(y\,|\,1)} \sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right| \\
&\quad\ge \mathrm{Re}\left\{\sum_{x\in\mathcal{X}} \hat{p}_a(1\,|\,x)w_x\right\} \\
&\qquad - \sum_{x'\in\mathcal{X}\setminus\{1\}} \left|\frac{\hat{q}_b(y\,|\,x')}{\hat{q}_b(y\,|\,1)}\right| \left|\sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right| \\
&\quad\ge 2^{-1}\tilde{\varepsilon}_\theta
\end{aligned} \tag{107}$$

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 (100), (101), (104), and (105), we obtain

$$\begin{aligned}
&|\,e^T\hat{R}_\eta(y)w\,| \\
&\quad\ge \left|\sum_{x'\in\mathcal{X}} q_\beta(y\,|\,x') \sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right| \\
&\qquad - \left|\sum_{x'\in\mathcal{X}} (\hat{q}_b(y\,|\,x') - q_\beta(y\,|\,x')) \sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right| \\
&\quad\ge \sum_{x'\in\mathcal{X}} q_\beta(y\,|\,x')\mathrm{Re}\left\{\sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right\} \\
&\qquad - \sum_{x'\in\mathcal{X}} |\hat{q}_b(y\,|\,x') - q_\beta(y\,|\,x')| \left|\sum_{x\in\mathcal{X}} \hat{p}_a(x'\,|\,x)w_x\right| \\
&\quad\ge 2^{-1}\tilde{\varepsilon}_\theta \sum_{x'\in\mathcal{X}} q_\beta(y\,|\,x') \\
&\quad\ge 2^{-1}\tilde{\varepsilon}_\theta q_\beta(y\,|\,1) \\
&\quad\ge 4^{-1}\tilde{\varepsilon}_\theta\,|\,\hat{q}_b(y\,|\,1)\,|
\end{aligned} \tag{108}$$

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 be concluded from (107) and (108) that $\hat{\phi}_\eta(w,y)$, $\hat{G}_\eta(w,y)$ are analytic 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, (102) and (106)–(108) imply

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

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) and ii) of the proposition, there exists a real number $\varepsilon_Q \in (0,1)$ such that $\varepsilon_Q \le p_\alpha(x'\,|\,x) \le \varepsilon_Q^{-1}$ 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

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

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 and K_θ are defined in the statement of Lemma 16). 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 16, 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 analytic 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., nonrecursive) maximum-likelihood estimators such as those studied in [6], [12], [27], and [36]. 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 nonanalytically parameterized hidden Markov models will be explored as well.

Appendix I

In this Appendix, the proofs of Lemmas 1–3 are provided.

Remark: Throughout the Appendix, the following convention is applied. Diacritic ˜ is used to denote a locally defined quantity, i.e., a quantity whose definition holds only in the proof where the quantity appears.

Lemma 17: Let Assumption 4 hold. Then, for any compact set $Q \subset \Theta$, there exist real numbers $\rho_Q \in (0,1)$, $L_{1,Q} \in [1,\infty)$ such that

$$\|F_\theta(u,V,y)\| \le L_{1,Q}\psi_Q(y)(1+\|V\|) \tag{109}$$

$$\|\nabla_\theta G_\theta(u,y)\| \le L_{1,Q} \tag{110}$$

$$\|H_\theta(u,V,y)\| \le L_{1,Q}(1+\|V\|) \tag{111}$$

$$\begin{aligned}&\|F_{\theta'}(u',V',y) - F_{\theta''}(u'',V'',y)\| \\ &\quad \le L_{1,Q}\psi_Q(y)(1+\|V'\|+\|V''\|) \\ &\qquad \cdot (\|\theta'-\theta''\|+\|u'-u''\|+\|V'-V''\|)\end{aligned} \tag{112}$$

$$\begin{aligned}&\|H_{\theta'}(u',V',y) - H_{\theta''}(u'',V'',y)\| \\ &\quad \le L_{1,Q}(1+\|V'\|+\|V''\|) \\ &\qquad \cdot (\|\theta'-\theta''\|+\|u'-u''\|+\|V'-V''\|)\end{aligned} \tag{113}$$

$$\begin{aligned}&|\hat{\phi}_{\eta'}(w',y) - \hat{\phi}_{\eta''}(w'',y)| \\ &\quad \le L_{1,Q}\psi_Q(y)(\|\eta'-\eta''\|+\|w'-w''\|)\end{aligned} \tag{114}$$

$$\begin{aligned}&|\hat{G}_{\eta'}(w',y) - \hat{G}_{\eta''}(w'',y)| \\ &\quad \le L_{1,Q}(\|\eta'-\eta''\|+\|w'-w''\|)\end{aligned} \tag{115}$$

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

Proof: Let $\rho_Q = \delta_Q/2$ (δ_Q is defined in Assumption 4). Then, Cauchy inequality for complex-analytic functions (see, e.g., [40, Prop. 2.1.3]) and Assumption 4 imply that there exists a real number $\tilde{L}_{1,Q} \in [1,\infty)$ such that

$$\max\left\{\left\|\nabla_{(\eta,w)}\hat{\phi}_\eta(w,y)\right\|, \left\|\nabla^2_{(\eta,w)}\hat{\phi}_\eta(w,y)\right\|\right\} \le \tilde{L}_{1,Q}\psi_Q(y)$$
$$\max\left\{\left\|\nabla_{(\eta,w)}\hat{G}^k_\eta(w,y)\right\|, \left\|\nabla^2_{(\eta,w)}\hat{G}^k_\eta(w,y)\right\|\right\} \le \tilde{L}_{1,Q}$$

for all $\eta \in V_{\rho_Q}(Q)$, $w \in V_{\rho_Q}(\mathcal{P}^{N_x})$, $y \in \mathcal{Y}$ [$\nabla_{(\eta,w)}$, $\nabla^2_{(\eta,w)}$ denote the gradient and Hessian with respect to (η, w), while $\hat{G}^k_\eta(w,y)$ stands for the kth component of $\hat{G}_\eta(w,y)$]. Consequently, there exists another real number $\tilde{L}_{2,Q} \in [1,\infty)$ such that

$$\begin{aligned}&\max\{\|\hat{\phi}_{\eta'}(w',y) - \hat{\phi}_{\eta''}(w'',y)\|, \\ &\qquad \|\nabla_w\hat{\phi}_{\eta'}(w',y) - \nabla_w\hat{\phi}_{\eta''}(w'',y)\|\} \\ &\qquad\quad \le \tilde{L}_{2,Q}\psi_Q(y)(\|\eta'-\eta''\|+\|w'-w''\|) \\ &\max\{\|\hat{G}_{\eta'}(w',y) - \hat{G}_{\eta''}(w'',y)\|, \\ &\qquad \|\nabla_w\hat{G}_{\eta'}(w',y) - \nabla_w\hat{G}_{\eta''}(w'',y)\|\} \\ &\qquad\quad \le \tilde{L}_{2,Q}(\|\eta'-\eta''\|+\|w'-w''\|)\end{aligned}$$

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

$$\begin{aligned}\|F_\theta(u,V,y)\| &\le \|\nabla_\theta\phi_\theta(u,y)\| + \|\nabla_u\phi_\theta(u,y)\|\|V\| \\ &\le \tilde{L}_{1,Q}\psi_Q(y)(1+\|V\|) \\ \|H_\theta(u,V,y)\| &\le \|\nabla_\theta G_\theta(u,y)\| + \|\nabla_u G_\theta(u,y)\|\|V\| \\ &\le \tilde{L}_{1,Q}N_x(1+\|V\|)\end{aligned}$$

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

$$\begin{aligned}&\|F_{\theta'}(u',V',y) - F_{\theta''}(u'',V'',y)\| \\ &\quad \le \|\nabla_\theta\phi_{\theta'}(u',y) - \nabla_\theta\phi_{\theta''}(u'',y)\| \\ &\qquad + \|\nabla_u\phi_{\theta'}(u',y) - \nabla_u\phi_{\theta''}(u'',y)\|\|V'\| \\ &\qquad + \|\nabla_u\phi_{\theta''}(u'',y)\|\|V'-V''\| \\ &\quad \le \tilde{L}_{2,Q}\psi_Q(y)(1+\|V'\|+\|V''\|)(\|\theta'-\theta''\|+\|u'-u''\|) \\ &\qquad + \tilde{L}_{1,Q}\psi_Q(y)\|V'-V''\|, \\ &\|H_{\theta'}(u',V',y) - H_{\theta''}(u'',V'',y)\| \\ &\quad \le \|\nabla_\theta G_{\theta'}(u',y) - \nabla_\theta G_{\theta''}(u'',y)\| \\ &\qquad + \|\nabla_u G_{\theta'}(u',y) - \nabla_u G_{\theta''}(u'',y)\|\|V'\| \\ &\qquad + \|\nabla_u G_{\theta''}(u'',y)\|\|V'-V''\| \\ &\quad \le \tilde{L}_{2,Q}(1+\|V'\|+\|V''\|)(\|\theta'-\theta''\|+\|u'-u''\|) \\ &\qquad + \tilde{L}_{1,Q}\|V'-V''\|\end{aligned}$$

for each $\theta', \theta'' \in Q$, $u', u'' \in \mathcal{P}^{N_x}$, $V', V'' \in \mathbb{R}^{d_\theta \times N_x}$. Then, it can be deduced that there exists a real number $L_{1,Q} \in [1, \infty)$ such that (109)–(115) hold for each $\theta, \theta', \theta'' \in Q$, $\eta', \eta'' \in V_{\rho_Q}(Q)$, $u, u', u'' \in \mathcal{P}^{N_x}$, $w', w'' \in V_{\rho_Q}(\mathcal{P}^{N_x})$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. $\square$

Proof of Lemma 2: For $\theta \in \Theta$, $u \in [0,\infty)^{N_x} \setminus \{0\}$, $y \in \mathcal{Y}$, let $A_\theta(u,y) = \nabla_u G_\theta(u,y)$, $B_\theta(u,y) = \nabla_\theta G_\theta(u,y)$. For $u \in [0,\infty)^{N_x} \setminus \{0\}$, $n > m \geq 0$ and sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in Θ, $\mathcal{Y}$ (respectively), let $A^{n:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u) = I$ (I denotes the $N_x \times N_x$ unit matrix) and

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

Then, it is easy to demonstrate

$$H^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u,V) = \sum_{i=m}^{n-1} B_{\vartheta_{i+1}}(G^{m:i}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u), y_{i+1}) A^{i+1:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}\left(G^{m:i+1}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)\right) + V A^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)$$

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 $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in Θ, $\mathcal{Y}$ (respectively). Let $Q \subset \Theta$ be an arbitrary compact set. Then, using [39, Th. 3.1, Lemma 6.6] (with a few straightforward modifications), it can be deduced from Assumption 3 that there exist real numbers $\varepsilon_{1,Q} \in (0,1)$, $\tilde{L}_Q \in [1,\infty)$ such that

$$\|G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(w') - G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(w'')\| \leq \tilde{L}_Q \varepsilon_{1,Q}^{n-m} \left\| \frac{w'}{e^T w'} - \frac{w''}{e^T w''} \right\| \tag{116}$$

$$\|A^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)\| \leq \tilde{L}_Q \varepsilon_{1,Q}^{n-m} \tag{117}$$

hold for all $u \in \mathcal{P}^{N_x}$, $w', w'' \in [0,\infty)^{N_x} \setminus \{0\}$, $n \geq m \geq 0$ and any sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in Q, $\mathcal{Y}$.[4] Consequently, we get

$$\begin{aligned} &\|G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(w') - G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(w'')\| \\ &\quad \leq \tilde{L}_Q \varepsilon_{1,Q} \frac{\|w' - w''\|(e^T w'') + \|w''\| \, |e^T(w' - w'')|}{(e^T w')(e^T w'')} \\ &\quad \leq 2(N_x + 1)\tilde{L}_Q \varepsilon_{1,Q} \|w' - w''\| \end{aligned}$$

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

Due to Lemma 17 and (117), we have

$$\begin{aligned} \|H^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u,V)\| &\leq \tilde{L}_Q \varepsilon_{1,Q}^{n-m} \|V\| + L_{1,Q} \tilde{L}_Q \sum_{i=m}^{n-1} \varepsilon_{1,Q}^{n-i-1} \\ &\leq \tilde{L}_Q \|V\| + L_{1,Q} \tilde{L}_Q (1 - \varepsilon_{1,Q})^{-1} \end{aligned}$$

for all $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq m \geq 0$ and any sequences $\boldsymbol{\vartheta} = \{\vartheta_n\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in Q, $\mathcal{Y}$. Then, it is clear that there exist real numbers $\varepsilon_{1,Q} \in (0,1)$, $C_{2,Q} \in [1,\infty)$ such that (13) and (14) hold for all $u \in \mathcal{P}^{N_x}$, $w', w'' \in \mathcal{Q}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$ and any sequence $\boldsymbol{\vartheta} = \{\vartheta\}_{n\geq 0}$, $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in Q, $\mathcal{Y}$. $\square$

[4]To deduce this, note that u, V, $y_{0:n}$, $G^{0:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)$, $A^{0:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)V$ have the same meaning, respectively, as quantities μ, $\tilde{\mu}$, y^n, $F^n_\theta(\mu, y^n)$, $\tilde{G}^n_\theta(\mu, \tilde{\mu}, y^n)$ appearing in [39]. Inequality (116) can also be obtained from [25, Th. 2.1] or [26, Th. 4.1]. Similarly, (117) can be deduced from [24, Lemmas 3.4 and 4.3] (notice that $G^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)$, $A^{m:n}_{\boldsymbol{\vartheta},\boldsymbol{y}}(u)$ have the same meaning, respectively, as $M_{m,n}$, $V[M_{m,n}, p_m]$ specified in [24, Sec. 5]).

Lemma 18: Let Assumptions 3 and 4 hold. Then, for any compact set $Q \subset \Theta$, there exists a real number $L_{2,Q} \in [1,\infty)$ such that

$$\|G^{0:n}_{\theta',\boldsymbol{y}}(u) - G^{0:n}_{\theta'',\boldsymbol{y}}(u)\| \leq L_{2,Q}\|\theta' - \theta''\| \tag{118}$$

$$\|H^{0:n}_{\theta',\boldsymbol{y}}(u,V) - H^{0:n}_{\theta'',\boldsymbol{y}}(u,V)\| \leq L_{2,Q}\|\theta' - \theta''\|(1 + \|V\|) \tag{119}$$

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 $\boldsymbol{y} = \{y_n\}_{n\geq 1}$ in $\mathcal{Y}$.

Proof: Let $Q \subset \Theta$ be arbitrary set. Then, using [39, Th. 3.2] (with a few obvious modifications), it can be deduced that there exist real numbers $\tilde{\varepsilon}_Q \in (0,1)$, $\tilde{L}_Q \in [1,\infty)$ such that

$$\begin{aligned} &\|H^{m:n}_{\theta,\boldsymbol{y}}(u',V') - H^{m:n}_{\theta,\boldsymbol{y}}(u'',V'')\| \\ &\leq \tilde{L}_Q \tilde{\varepsilon}_Q^{n-m}(\|u' - u''\|(1 + \|V'\| + \|V''\|) + \|V' - V''\|) \end{aligned} \tag{120}$$

for all $\theta \in Q$, $u', u'' \in \mathcal{P}^{N_x}$, $V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq m \geq 0$ and any sequence $\boldsymbol{y} = \{y_n\}_{n\geq 0}$ in $\mathcal{Y}$.

Let $\boldsymbol{y} = \{y_n\}_{n\geq 0}$ be an arbitrary sequence in $\mathcal{Y}$. It is straightforward to verify

$$\begin{aligned} &G^{0:n}_{\theta',\boldsymbol{y}}(u) - G^{0:n}_{\theta'',\boldsymbol{y}}(u) \\ &\quad = \sum_{i=0}^{n-1} \left(G^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) - G^{i+1:n}_{\theta',\boldsymbol{y}}(G^{0:i+1}_{\theta'',\boldsymbol{y}}(u)) \right) \end{aligned} \tag{121}$$

$$\begin{aligned} &H^{0:n}_{\theta',\boldsymbol{y}}(u,V) - H^{0:n}_{\theta'',\boldsymbol{y}}(u,V) \\ &\quad = \sum_{i=0}^{n-1} \left(H^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)) \right. \\ &\qquad \left. - H^{i+1:n}_{\theta',\boldsymbol{y}}(G^{0:i+1}_{\theta'',\boldsymbol{y}}(u), H^{0:i+1}_{\theta'',\boldsymbol{y}}(u,V)) \right) \end{aligned} \tag{122}$$

for all $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq 0$. Since

$$G^{0:i+1}_{\theta'',\boldsymbol{y}}(u) = G^{i:i+1}_{\theta'',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u))$$

$$G^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) = G^{i+1:n}_{\theta',\boldsymbol{y}}\left(G^{i:i+1}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u))\right)$$

for any $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $0 \leq i < n$ Lemmas 2 and 17 yield

$$\begin{aligned} &\|G^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) - G^{i+1:n}_{\theta',\boldsymbol{y}}(G^{0:i+1}_{\theta'',\boldsymbol{y}}(u))\| \\ &\quad \leq C_{2,Q} \varepsilon_{1,Q}^{n-i-1} \left\| G^{i:i+1}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) - G^{i:i+1}_{\theta'',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) \right\| \\ &\quad \leq C_{2,Q} L_{1,Q} \varepsilon_{1,Q}^{n-i-1} \|\theta' - \theta''\| \end{aligned} \tag{123}$$

for the same θ', θ'', u, n, i [notice that $G^{i:i+1}_{\theta,\boldsymbol{y}}(u) = G_\theta(u, y_{i+1})$]. On the other side, as

$$\begin{aligned} &H^{0:i+1}_{\theta'',\boldsymbol{y}}(u,V) \\ &\quad = H^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right) \\ &H^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)) \\ &\quad = H^{i+1:n}_{\theta',\boldsymbol{y}}\left(G^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)\right), \right. \\ &\qquad \left. H^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right) \end{aligned}$$

for each $\theta', \theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $0 \leq i < n$, Lemmas 2 and 17 and (120) imply

$$\begin{aligned}
&\|H^{i:n}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)) \\
&\quad - H^{i+1:n}_{\theta',\boldsymbol{y}}(G^{0:i+1}_{\theta'',\boldsymbol{y}}(u), H^{0:i+1}_{\theta'',\boldsymbol{y}}(u,V))\| \\
&\quad \leq \tilde{L}_Q \tilde{\varepsilon}_Q^{n-i-1} \left\| G^{i:i+1}_{\theta',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) - G^{i:i+1}_{\theta'',\boldsymbol{y}}(G^{0:i}_{\theta'',\boldsymbol{y}}(u)) \right\| \\
&\qquad \cdot \left(1 + \left\| H^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right\| \right. \\
&\qquad\quad \left. + \left\| H^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right)\right\| \right) \\
&\qquad + \tilde{L}_Q \tilde{\varepsilon}_Q^{n-i-1} \left\| H^{i:i+1}_{\theta',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right) \right. \\
&\qquad\qquad \left. - H^{i:i+1}_{\theta'',\boldsymbol{y}}\left(G^{0:i}_{\theta'',\boldsymbol{y}}(u)), H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\right) \right\| \\
&\quad \leq (3L_{1,Q}+1) L_{1,Q} \tilde{L}_Q \tilde{\varepsilon}_Q^{n-i-1} \|\theta' - \theta''\| \\
&\qquad \cdot (1 + \|H^{0:i}_{\theta'',\boldsymbol{y}}(u,V)\|) \\
&\quad \leq 8 C_{2,Q} L_{1,Q}^2 \tilde{L}_Q \tilde{\varepsilon}_Q^{n-i-1} \|\theta' - \theta''\| (1 + \|V\|)
\end{aligned} \tag{124}$$

for the same θ', θ'', u, V, n, i [notice that $H^{i:i+1}_{\theta,\boldsymbol{y}}(u,V) = H_\theta(u,V,y_{i+1})$]. Combining (121)–(124), we conclude that there exists a real number $L_{2,Q} \in [1,\infty)$ such that (118) and (119) 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 $\boldsymbol{y} = \{y_n\}_{n \geq 1}$ in $\mathcal{Y}$. □

Proof of Lemma 3: Let $\boldsymbol{Y} = \{Y_n\}_{n \geq 0}$. Moreover, for $\theta \in \Theta$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$, let

$$\begin{aligned}
\tilde{\phi}_\theta(u,x) &= E(\phi_\theta(u,Y_1) \mid X_0 = x) \\
\tilde{F}_\theta(u,V,x) &= E(F_\theta(u,V,Y_1) \mid X_0 = x)
\end{aligned}$$

while

$$\begin{aligned}
&\tilde{\phi}^n_\theta(u,x,y) \\
&\quad = E\left(\tilde{\phi}_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), X_n) \middle| X_0 = x, Y_0 = y \right) \\
&\tilde{F}^n_\theta(u,V,x,y) \\
&\quad = E\left(\tilde{F}_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), H^{0:n}_{\theta,\boldsymbol{Y}}(u,V), X_n) \middle| X_0 = x, Y_0 = y \right)
\end{aligned}$$

for $n \geq 0$. Consequently, Lemma 17 yields

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

for all $\theta \in Q$, $u, u', u'' \in \mathcal{P}^{N_x}$, $V, V', V'' \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$ and any compact set $Q \subset \Theta$, where $\tilde{L}_{1,Q} = L_{1,Q} \max_{x' \in \mathcal{X}} \int \psi_Q(y) Q(dy \mid x')$. Then, owing to [39, Th. 4.1 and 4.2], there exist functions $\psi: \Theta \to \mathbb{R}$, $g: \Theta \to \mathbb{R}^{d_\theta}$ and for any compact set $Q \subset \Theta$, there exist real numbers $\varepsilon_{2,Q} \in (0,1)$, $\tilde{L}_{2,Q} \in [1,\infty)$ such that the following is true:

$$|\tilde{\phi}^n_\theta(u,x,y) - \psi(\theta)| \leq \tilde{L}_{2,Q} \varepsilon^n_{2,Q} \tag{125}$$
$$\|\tilde{F}^n_\theta(u,V,x,y) - g(\theta)\| \leq \tilde{L}_{2,Q} \varepsilon^n_{2,Q}(1 + \|V\|^2) \tag{126}$$

for all $\theta \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $n \geq 0$.[5] On the other side, it is easy to show

$$\begin{aligned}
(\Pi^n \phi)(\theta,\zeta) &= E\left(\phi_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), Y_{n+1}) \mid X_1 = x, Y_1 = y\right) \\
&= \tilde{\phi}^{n-1}_\theta(G_\theta(u,y),x,y)
\end{aligned}$$

for all $\theta \in \Theta$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $n \geq 1$. It is also easy to demonstrate

$$\begin{aligned}
&(P^n F)(\theta,z) \\
&\quad = E\left(F_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), H^{0:n}_{\theta,\boldsymbol{Y}}(u,V), Y_{n+1}) \mid X_1 = x, Y_1 = y\right) \\
&\quad = \tilde{F}^{n-1}_\theta(G_\theta(u,y), H_\theta(u,V,y), x, y)
\end{aligned}$$

for all $\theta \in \Theta$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $z = (x,y,u,V) \in \mathcal{S}_z$, $n \geq 1$. Thus, for any compact set $Q \subset \Theta$, there exists a real number $\tilde{L}_{3,Q} \in [1,\infty)$ such that

$$|(\Pi^n \phi)(\theta,\zeta) - \psi(\theta)| \leq \tilde{L}_{3,Q} \varepsilon^n_{2,Q} \tag{127}$$
$$\|(P^n F)(\theta,z) - g(\theta)\| \leq \tilde{L}_{3,Q} \varepsilon^n_{2,Q}(1 + \|V\|^2) \tag{128}$$

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

It is straightforward to verify

$$\begin{aligned}
&E\left(\frac{\log p^n_\theta(Y_1,\ldots,Y_{n+1})}{n} \middle| X_1 = x, Y_1 = y \right) \\
&\quad = E\left(\frac{1}{n} \sum_{i=0}^{n-1} \phi_\theta(G^{0:i}_{\theta,\boldsymbol{Y}}(u_\theta), Y_{i+1}) \middle| X_1 = x, Y_1 = y \right) \\
&\quad = \frac{1}{n} \sum_{i=1}^{n-1} \tilde{\phi}^{i-1}_\theta(G_\theta(u_\theta,y),x,y) + \frac{\phi_\theta(u_\theta,Y_1)}{n+1}
\end{aligned} \tag{129}$$

for each $\theta \in \Theta$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $n > 1$, where $u_\theta = [P(X^\theta_1 = 1) \cdots P(X^\theta_1 = N_x)]^T$. It is also easy to demonstrate

$$\nabla_\theta \left(\phi_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), Y_{n+1})\right) = F_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), H^{0:n}_{\theta,\boldsymbol{Y}}(u,0), Y_{n+1}) \tag{130}$$

for all $\theta \in \Theta$, $u \in \mathcal{P}^{N_x}$, $n \geq 1$ (here, 0 stands for the $d_\theta \times N_x$ zero matrix). Combining (125) and (129), we deduce that on Θ, $f(\cdot)$ is well defined and satisfies $\psi(\cdot) = f(\cdot)$. On the other side, Lemmas 2 and 17 yield

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

for each $\theta \in Q$, $u \in \mathcal{P}^{N_x}$, $n \geq 1$ and any compact set $Q \subset \Theta$. Therefore

$$\begin{aligned}
&E\left(\|F_\theta(G^{0:n}_{\theta,\boldsymbol{Y}}(u), H^{0:n}_{\theta,\boldsymbol{Y}}(u,0), Y_{n+1})\| \middle| X_1 = x, Y_1 = y\right) \\
&\quad \leq 2 C_{1,Q} \tilde{L}_{1,Q}(1 + \|V\|) \\
&\quad < \infty
\end{aligned}$$

[5] The same result can also be obtained from [23, Th. 5.4]

for all $\theta \in Q$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $n \geq 1$ and any compact set $Q \subset \Theta$. Consequently, the dominated convergence theorem and (130) imply

$$\begin{aligned}
&\nabla_\theta(\Pi^n\phi)(\theta,\zeta)\\
&\quad= E\left(\nabla_\theta\left(\phi_\theta(G^{0:n}_{\theta,\mathbf{Y}}(u),Y_{n+1})\right)\middle|\,X_1=x,Y_1=y\right)\\
&\quad= E\left(F_\theta(G^{0:n}_{\theta,\mathbf{Y}}(u),H^{0:n}_{\theta,\mathbf{Y}}(u,0),Y_{n+1})\middle|\,X_1=x,Y_1=y\right)\\
&\quad= (P^nF)(\theta,(x,y,u,0))
\end{aligned}$$

for all $\theta \in \Theta$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $n \geq 1$. 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 all $\zeta \in \mathcal{S}_\zeta$, $z \in \mathcal{S}_z$ and any compact set $Q \subset \Theta$ [due to (127) and (128)], we conclude that on Θ, $f(\cdot)$ is differentiable and satisfies $g(\cdot) = \nabla\psi(\cdot) = \nabla f(\cdot)$. Owing to Lemmas 2, 17, and 18, we have

$$\begin{aligned}
&\|F_{\theta'}(G^{0:n}_{\theta',\mathbf{Y}}(u),H^{0:n}_{\theta',\mathbf{Y}}(u,V),Y_{n+1})\\
&\quad- F_{\theta''}(G^{0:n}_{\theta'',\mathbf{Y}}(u),H^{0:n}_{\theta'',\mathbf{Y}}(u,V),Y_{n+1})\|\\
&\qquad\leq L_{1,Q}\psi_Q(Y_{n+1})(1+\|H^{0:n}_{\theta',\mathbf{Y}}(u,V)\|+\|H^{0:n}_{\theta'',\mathbf{Y}}(u,V)\|)\\
&\qquad\quad\cdot\left(\|\theta'-\theta''\|+\|G^{0:n}_{\theta',\mathbf{Y}}(u)-G^{0:n}_{\theta'',\mathbf{Y}}(u)\|\right.\\
&\qquad\qquad\left.+\|H^{0:n}_{\theta',\mathbf{Y}}(u,V)-H^{0:n}_{\theta'',\mathbf{Y}}(u,V)\|\right)\\
&\qquad\leq 9C_{1,Q}L_{1,Q}L_{2,Q}\psi_Q(Y_{n+1})(1+\|V\|)^2\|\theta'-\theta''\|
\end{aligned}$$

for all $\theta',\theta'' \in Q$, $u \in \mathcal{P}^{N_x}$, $V \in \mathbb{R}^{d_\theta \times N_x}$, $n \geq 1$. Therefore

$$\begin{aligned}
&\|(P^nF)(\theta',z)-(P^nF)(\theta'',z)\|\\
&\qquad\leq 9C_{1,Q}L_{2,Q}\tilde{L}_{1,Q}\|\theta'-\theta''\|(1+\|V\|)^2 \qquad (131)
\end{aligned}$$

for each $\theta',\theta'' \in Q, z = (x,y,u,V) \in \mathcal{S}_z$. Since $\psi(\cdot) = f(\cdot)$ and $g(\cdot) = \nabla f(\cdot)$, it can be deduced from (127), (128), and (131) that for any compact set $Q \subset \Theta$, there exist real numbers $\varepsilon_{2,Q} \in (0,1), C_{3,Q} \in [1,\infty)$ such that (15)–(17) hold for all $\theta,\theta',\theta'' \in Q$, $\zeta = (x,y,u) \in \mathcal{S}_\zeta$, $z = (x,y,u,V) \in \mathcal{S}_z$, $n \geq 1$. □

Appendix II

Lemmas 5 and 6 are proved in this Appendix.

Lemma 19: 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

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

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

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

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

Proof of Lemma 5: Let $Q \subseteq \Theta$ be an arbitrary compact set. Owing to Lemmas 1 and 3, there exists a real number $\tilde{C}_{1,Q} \in [1,\infty)$ such that

$$\sum_{k=0}^\infty \|(P^kF)(\theta,z)-\nabla f(\theta)\| \leq \tilde{C}_{1,Q}\psi_Q(y)(1+\|V\|^2) \quad (132)$$

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

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

for each $\theta',\theta'' \in \Theta$, $z \in \mathcal{S}_z$, $n \geq 1$. Then, using Lemma 3, it can be deduced that there exist real numbers $\tilde{\varepsilon}_Q \in (0,1)$, $\tilde{C}_{2,Q} \in [1,\infty)$ such that

$$\begin{aligned}
&\left\|\sum_{k=1}^\infty((P^kF)(\theta',z)-\nabla f(\theta'))-\sum_{k=1}^\infty((P^kF)(\theta'',z)-\nabla f(\theta''))\right\|\\
&\quad\leq \tilde{C}_{2,Q}(1+\|V\|^2)(\tilde{\varepsilon}_Q^n+n\|\theta'-\theta''\|) \qquad (133)
\end{aligned}$$

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

$$N_{Q,s}(t)+\tilde{\varepsilon}_Q^{N_{Q,s}(t)} \leq \tilde{K}_{Q,s}t^s \qquad (134)$$

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

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

$$\begin{aligned}
\Phi(\theta,z) &= \sum_{k=0}^\infty((P^kF)(\theta,z)-\nabla f(\theta))\\
\varphi_{Q,s}(z) &= \tilde{C}_Q\tilde{K}_{Q,s}\psi_Q(y)(1+\|V\|^2).
\end{aligned}$$

Since

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

for all $\theta \in \Theta$, $z = (x, y, u, V) \in \mathcal{S}_z$, we deduce from (132) that $\Phi(\cdot, \cdot)$ is well defined, integrable, and satisfies (34) and (35) [notice that $(P\Phi)(\theta, z) = \sum_{k=1}^{\infty}((P^k F)(\theta, z) - \nabla f(\theta))$]. On the other hand, (133) and (134) 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 (133)]. Thus, (36) 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 2, we have

$$\begin{aligned} &\varphi_{Q,s}(Z_{n+1}) I_{\{\tau_Q > n\}} \\ &\quad = \tilde{C}_Q \tilde{K}_{Q,s} \psi_Q(Y_{n+1})(1 + \|H^{0:n}_{\boldsymbol{\theta},\boldsymbol{Y}}(U_0, V_0)\|^2) I_{\{\tau_Q > n\}} \\ &\quad \leq 4\tilde{C}_Q \tilde{K}_{Q,s} C^2_{4,Q} \psi_Q(Y_{n+1})(1 + \|V_0\|^2) \end{aligned} \tag{135}$$

for each $n \geq 0$ [notice that $H^{0:n}_{\boldsymbol{\theta},\boldsymbol{Y}}(U_0, V_0)$ depends only on the first n elements of $\boldsymbol{\theta}$, and that $\theta_1, \ldots, \theta_n \in Q$ is sufficient for (135) to hold]. Consequently

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

for all $z = (x, y, u, V) \in \mathcal{S}_z$, $n \geq 0$. Hence, (37) is true for all $z \in \mathcal{S}_z$. □

Proof of Lemma 6: Let $Q \subset \Theta$ be an arbitrary compact set, while t is an arbitrary number in $[0, r]$. 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 $\sum_{n=0}^{\infty} \alpha_n^{1+s} \gamma_n^r < \infty$ (its existence is demonstrated in Lemma 19, Appendix II), while

$$\begin{aligned} \tilde{C}_Q = \max\Bigg\{ &\|\nabla\Psi(\theta)\|, \frac{\|\Psi(\theta') - \Psi(\theta'')\|}{\|\theta' - \theta''\|^s}, \\ &\frac{\|\nabla f(\theta') - \nabla f(\theta'')\|}{\|\theta' - \theta''\|} : \theta, \theta', \theta'' \in Q \Bigg\}. \end{aligned}$$

Moreover, for $n \geq 1$, let

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

Then, it is straightforward to verify

$$\sum_{i=1}^{n} \alpha_i \gamma_i^t \Psi(\theta_i) \xi_i = \sum_{i=1}^{n} \alpha_i \gamma_i^t \psi_{1,i} + \sum_{i=1}^{n} \alpha_i \gamma_i^t \psi_{2,i} \tag{136}$$
$$+ \sum_{i=0}^{n-1} (\alpha_{i+1}\gamma_{i+1}^t - \alpha_i \gamma_i^t)\psi_{3,i} \tag{137}$$
$$- \alpha_n \gamma_n^t \psi_{3,n} + \alpha_0 \gamma_0^t \psi_{3,0} \tag{138}$$

for $n \geq 1$.

Owing to Assumption 1, we have

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

as $n \to \infty$. Consequently

$$\begin{aligned} \sum_{n=0}^{\infty} \alpha_n^s \alpha_{n+1} \gamma_{n+1}^t &= \sum_{n=0}^{\infty} (\alpha_n/\alpha_{n+1})^s \alpha_{n+1}^s \gamma_{n+1}^t \\ &< \infty \end{aligned} \tag{139}$$

$$\begin{aligned} \sum_{n=0}^{\infty} |\,\alpha_n \gamma_n^t - \alpha_{n+1}\gamma_{n+1}^t\,| &\leq \sum_{n=0}^{\infty} \alpha_n \,|\,\gamma_n^t - \gamma_{n+1}^t\,| \\ &\quad + \sum_{n=0}^{\infty} |\,\alpha_n - \alpha_{n+1}\,|\,\gamma_{n+1}^t \\ &< \infty. \end{aligned} \tag{140}$$

On the other side, as a consequence of Lemma 5, we get

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

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

$$\begin{aligned} E_{\theta,z}(|\,\psi_{3,n}\,|^2 I_{\{\tau_Q > n\}}) &\leq \tilde{C}_Q^2 E_{\theta,z}(\varphi^2_{Q,s}(Z_{n+1}) I_{\{\tau_Q > n\}}) \\ E_{\theta,z}\left(|\,\phi''_n\,| I_{\{\tau_Q > n\}}\right) &\leq \tilde{C}_Q E_{\theta,z}(\|\theta_{n+1} - \theta_n\|^2 I_{\{\tau_Q > n\}}) \\ &\leq \tilde{C}_Q \alpha_n^2 E_{\theta,z}(\varphi^2_{Q,s}(Z_{n+1}) I_{\{\tau_Q > n\}}) \end{aligned}$$

for all $\theta \in \Theta$, $z \in \mathcal{S}_z$, $n \geq 1$. Then, Lemma 5 and (139) yield

$$\begin{aligned} &E_{\theta,z}\left(\sum_{n=1}^{\infty} \alpha_n^2 \gamma_n^{2t} |\,\psi_{1,n}\,|^2 I_{\{\tau_Q > n\}}\right) \\ &\quad \leq 4\tilde{C}_Q^2 \left(\sum_{n=1}^{\infty} \alpha_n^2 \gamma_n^{2t}\right) \sup_{n \geq 0} E_{\theta,z}\left(\varphi^2_{Q,s}(Z_{n+1}) I_{\{\tau_Q > n\}}\right) \\ &\quad < \infty \end{aligned} \tag{141}$$

$$E_{\theta,z}\left(\sum_{n=1}^{\infty}\alpha_n\gamma_n^t\,|\,\psi_{2,n}\,|\,I_{\{\tau_Q>n\}}\right)$$
$$\leq 2\tilde{C}_Q\left(\sum_{n=1}^{\infty}\alpha_{n-1}^s\alpha_n\gamma_n^t\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right)$$
$$<\infty \tag{142}$$

for any $\theta\in\Theta$, $z\in\mathcal{S}_z$. Moreover, Lemma 5 and (140) imply

$$E_{\theta,z}\left(\sum_{n=1}^{\infty}|\,\alpha_n\gamma_n^t-\alpha_{n+1}\gamma_{n+1}^t\,|\;|\,\psi_{3,n}\,|\,I_{\{\tau_Q>n\}}\right)$$
$$\leq \tilde{C}_Q\left(\sum_{n=1}^{\infty}|\,\alpha_n\gamma_n^t-\alpha_{n+1}\gamma_{n+1}^t\,|\right)\cdot\sup_{n\geq 0}\left(E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right)\right)^{1/2}$$
$$<\infty \tag{143}$$

$$E_{\theta,z}\left(\sum_{n=1}^{\infty}\alpha_{n+1}^2\gamma_{n+1}^{2t}\,|\,\psi_{3,n}\,|^2I_{\{\tau_Q>n\}}\right)$$
$$\leq \tilde{C}_Q^2\left(\sum_{n=1}^{\infty}\alpha_{n+1}^2\gamma_{n+1}^{2t}\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right)$$
$$<\infty \tag{144}$$

$$E_{\theta,z}\left(\sum_{n=0}^{\infty}|\,\phi_n''\,|\,I_{\{\tau_Q>n\}}\right)$$
$$\leq \tilde{C}_Q\left(\sum_{n=0}^{\infty}\alpha_n^2\right)\sup_{n\geq 0}E_{\theta,z}\left(\varphi_{Q,s}^2(Z_{n+1})I_{\{\tau_Q>n\}}\right)$$
$$<\infty \tag{145}$$

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

Owing to (142)–(145), series

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

converge w.p.1 on $\bigcap_{n=0}^{\infty}\{\theta_n\in Q\}$ and $\lim_{n\to\infty}\alpha_n\gamma_n^t\psi_{3,n}=0$ w.p.1 on the same event. On the other hand, we have

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

w.p.1 for every $\theta\in\Theta$, $z\in\mathcal{S}_z$, $n\geq 1$, where $\mathcal{F}_n=\sigma\{\theta_0,Z_0,\ldots,\theta_n,Z_n\}$. Hence, $\left\{\psi_{1,n}I_{\{\tau_Q>n\}}\right\}_{n\geq 0}$ is a martingale-difference sequence. Combining this with (141) and using the martingale convergence theorem (see [14, Corollary 2.2]), we get that $\sum_{n=1}^{\infty}\alpha_n\gamma_n^t\psi_{1,n}$ is convergent w.p.1 on $\bigcap_{n=0}^{\infty}\{\theta_n\in Q\}$. Then, (136) implies that $\sum_{n=0}^{\infty}\alpha_n\gamma_n^t\Psi(\theta_n)\xi_n$ converges w.p.1 on $\bigcap_{n=0}^{\infty}\{\theta_n\in Q\}$. Hence, event N_0 with the required properties exits. □

APPENDIX III

In this Appendix, we demonstrate that Assumption 4 does not hold for the models specified in Section III-D when $\mathcal{B}=(0,\infty)^{N_x}\times\mathbb{R}^{N_x}$ (i.e., when Θ includes points from $\mathcal{A}\times\tilde{\mathcal{B}}$). To do so, we use the following notation. For $a\in\mathbb{C}^{d_\alpha}$, $b=[l_1\cdots l_{N_x}m_1\cdots m_{N_x}]^T\in\mathbb{C}^{2N_x}$, $w=[w_1\cdots w_{N_x}]^T\in\mathbb{C}^{N_x}$, $x'\in\mathcal{X}$, $y\in\mathcal{Y}$, and $\eta=[a^Tb^T]^T$, let

$$\hat{h}_{x'}(\eta,w,y)=\sum_{x\in\mathcal{X}}\sqrt{l_{x'}}\exp(-l_{x'}(y-m_{x'})^2)\hat{p}_\alpha(x'\,|\,x)w_x$$
$$\hat{g}_{x'}(\eta,w,y)=\frac{\hat{h}_{x'}(\eta,w,y)}{\sum_{x''\in\mathcal{X}}\hat{h}_{x''}(\eta,w,y)}.$$

Remark: $[\hat{g}_1(\eta,w,y)\cdots\hat{g}_{N_x}(\eta,w,y)]^T$ would be a unique analytic continuation of $G_\theta(u,y)$ if such a continuation existed. However, due to Lemma 20, even if the continuation exists, it cannot satisfy (4) (i.e., Assumption 4).

Lemma 20: Let conditions i) and ii) of Proposition 4 hold, while $\theta\in\mathcal{A}\times\tilde{\mathcal{B}}$, $u\in\mathcal{P}^{N_x}$ are arbitrary vectors. Then, there exist sequences $\{\eta_n\}_{n\geq 0}$, $\{y_n\}_{n\geq 0}$ in $\mathbb{C}^{d_\theta}$, $\mathcal{Y}$ (respectively) such that $\lim_{n\to\infty}\eta_n=\theta$ and

$$\lim_{n\to\infty}\sum_{x\in\mathcal{X}}|\,\hat{g}_x(\eta_n,u,y_n)\,|=\infty \tag{146}$$

($\{\eta_n\}_{n\geq 0}$, $\{y_n\}_{n\geq 0}$ may be dependent on θ).

Proof: θ admits the representation $\theta=[\alpha^T\beta^T]^T$, where $\alpha\in\mathcal{A}$, $\beta=[\lambda_1\cdots\lambda_{N_x}\mu_1\cdots\mu_{N_x}]^T\in\tilde{\mathcal{B}}$. Without loss of generality, it can be assumed that there exists an integer $\tilde{N}$ (which may depend on θ) with the following properties: i) $2\leq\tilde{N}\leq N_x$, ii) $\lambda_x=\lambda$ for each $x\in\tilde{\mathcal{X}}$, iii) $\lambda_x>\lambda$ for all $x\in\mathcal{X}\setminus\tilde{\mathcal{X}}$, where $\lambda=\min_{x\in\mathcal{X}}\lambda_x$, $\tilde{\mathcal{X}}=\{1,\ldots,\tilde{N}\}$. It can also be assumed without loss of generality that $\mu_1=\mu$, where $\mu=\max_{x\in\tilde{\mathcal{X}}}\mu_x$.

Let $r_x=\sum_{x'\in\mathcal{X}}p_\alpha(x\,|\,x')u_{x'}$ for $x\in\mathcal{X}$, while $s=\sum_{x\in\tilde{\mathcal{X}}\setminus\{1\}}r_x/r_1$, $t=(\pi^2+\log^2 s)^{1/2}$, $\varphi=\pi-\arctan(\pi/t)$. Moreover, let $y_n=\mu_1+n$, $\rho_n=t/n^2$ for $n\geq 1$, while $l_{1,n}=\lambda+\rho_ne^{i\varphi}$, $l_{x',n}=\lambda n^2/(\mu_1-\mu_{x'}+n)^2$, $l_{x'',n}=\lambda_{x''}$ for $x'\in\tilde{\mathcal{X}}\setminus\{1\}$, $x''\in\mathcal{X}\setminus\tilde{\mathcal{X}}$, $n\geq 1$. Furthermore, let $b_n=[l_{1,n}\cdots l_{N_x,n}\mu_1\cdots\mu_{N_x}]^T$, $\eta_n=[\alpha^Tb_n^T]^T$ for $n\geq 1$. Then, it is straightforward to verify

$$\hat{h}_1(\eta_n,y_n,u)=\sqrt{\lambda+\rho_ne^{i\varphi}}\exp(-\lambda n^2)\exp(-te^{i\varphi})r_1$$
$$=-\sqrt{\lambda+\rho_ne^{i\varphi}}\exp(-\lambda n^2)r_1s.$$

It is also easy to show

$$\sum_{x\in\mathcal{X}\setminus\{1\}}\hat{h}_x(\eta_n,y_n,u)$$
$$=\sum_{x\in\tilde{\mathcal{X}}\setminus\{1\}}\frac{\sqrt{\lambda}n}{\mu_1-\mu_x+n}\exp(-\lambda n^2)r_x$$
$$+\sum_{x\in\mathcal{X}\setminus\tilde{\mathcal{X}}}\sqrt{\lambda_x}\exp(-\lambda_x(\mu_1-\mu_x+n)^2)$$

[notice that $e^{i\varphi}=(-\log s+i\pi)/t$]. Consequently

$$\lim_{n\to\infty}\exp(\lambda n^2)\hat{h}_1(\eta_n,y_n,u)=-\sqrt{\lambda}r_1s\neq 0$$
$$\lim_{n\to\infty}\exp(\lambda n^2)\sum_{x\in\mathcal{X}}\hat{h}_x(\eta_n,y_n,u)$$
$$=-\sqrt{\lambda}r_1s+\sqrt{\lambda}\sum_{x\in\tilde{\mathcal{X}}\setminus\{1\}}r_x=0$$

[notice that $\lim_{n\to\infty}\exp(\lambda n^2)\exp(-\lambda_x(\mu_1-\mu_x+n)^2)=0$ for all $x\in\mathcal{X}\setminus\tilde{\mathcal{X}}$; also notice that due to condition ii) of Proposition 2, $r_x>0$ for each $x\in\mathcal{X}$], wherefrom (146) directly follows. □

ACKNOWLEDGMENT

The author would like to thank the associate editor and the anonymous reviewers for valuable comments which significantly improved the original manuscript.

REFERENCES

[1] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statistics*, vol. 37, pp. 1554–1563, 1966.

[2] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.

[3] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[4] D. P. Bertsekas and J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," *SIAM J. Optim.*, vol. 10, pp. 627–642, 2000.

[5] P. J. Bickel and Y. Ritov, "Inference in hidden Markov models I. Local asymptotic normality in the stationary case," *Bernoulli*, vol. 2, pp. 199–228, 1996.

[6] P. J. Bickel, Y. Ritov, and T. Ryden, "Asymptotic normality of the maximum likelihood estimator for general hidden Markov models," *Ann. Stat.*, vol. 26, pp. 1614–1635, 1998.

[7] V. S. Borkar and S. P. Meyn, "The ODE method for convergence of stochastic approximation and reinforcement learning," *SIAM J. Control Optim.*, vol. 38, pp. 447–469, 2000.

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

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

[10] P. Del Moral and A. Guionnet, "On the stability of interacting processes with applications to filtering and genetic algorithms," *Annales de l'Institut Henri Poincaré, Probabilits et Statistique*, vol. 37, pp. 155–194, 2001.

[11] R. Douc and C. Matias, "Asymptotics of the maximum likelihood estimator for hidden Markov models for general hidden Markov models," *Bernoulli*, vol. 7, pp. 381–420, 2002.

[12] R. Douc, E. Moulines, and T. Ryden, "Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime," *Ann. Stat.*, vol. 32, pp. 2254–2304, 2004.

[13] Y. Ephraim and N. Merhav, "Hidden Markov models," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2008.

[14] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Applications*. New York: Academic, 1980.

[15] G. Han and B. Marcus, "Analyticity of entropy rate of hidden Markov chains," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5251–5266, Dec. 2006.

[16] G. Han and B. Marcus, "Derivatives of entropy rate in special families of hidden Markov chains," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2642–2652, Jul. 2007.

[17] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state channels based on Lyapunov exponents of random matrices," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3509–3532, Aug. 2006.

[18] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*. Cambridge, MA: Birkhäuser, 2002.

[19] V. Krishnamurthy and G. G. Yin, "Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 458–476, Feb. 2002.

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

[21] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. New York: Springer-Verlag, 2003.

[22] F. Le Gland and L. Mével, "Recursive identification of HMM's with observations in a finite set," in *Proc. 34th Conf. Decision Control*, 1995, pp. 216–221.

[23] F. Le Gland and L. Mével, "Recursive estimation in hidden Markov models," in *Proc. 36th Conf. Decision Control*, 1997, pp. 3468–3473.

[24] F. Le Gland and L. Mével, "Basic properties of the projective product with application to products of column-allowable nonnegative matrices," *Math. Control Signals Syst.*, vol. 13, pp. 41–62, 2000.

[25] F. Le Gland and L. Mével, "Exponential forgetting and geometric ergodicity in hidden Markov models," *Math. Control Signals Syst.*, vol. 13, pp. 63–93, 2000.

[26] F. Le Gland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters," *Ann. Appl. Probab.*, vol. 14, pp. 144–187, 2004.

[27] B. G. Leroux, "Maximum-likelihood estimation for hidden Markov models," *Stochastic Processes Their Appl.*, vol. 40, pp. 127–143, 1992.

[28] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[29] S. Lojasiewicz, "Sur le probléme de la division," *Studia Mathematica*, vol. 18, pp. 87–136, 1995.

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

[31] L. Mevel and L. Finesso, "Asymptotical statistics of misspecified hidden Markov models," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1123–1132, Jul. 2004.

[32] E. Ordentlich and T. Weissman, "On the optimality of symbol-by-symbol filtering and denoising," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 19–40, Jan. 2006.

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

[34] B. T. Polyak and Y. Z. Tsypkin, "Criterion algorithms of stochastic optimization," *Autom. Remote Control*, vol. 45, pp. 766–774, 1984.

[35] B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software, 1987.

[36] T. Ryden, "Consistent and asymptotically normal parameter estimates for hidden Markov models," *Ann. Stat.*, vol. 22, pp. 1884–1895, 1994.

[37] T. Ryden, "On recursive estimation for hidden Markov models," *Stochastic Processes Their Appl.*, vol. 66, pp. 79–96, 1997.

[38] A. Schönhuth, "On analytic properties of entropy rate," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2119–2127, May 2009.

[39] V. B. Tadić and A. Doucet, "Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models," *Stochastic Processes Their Appl.*, vol. 115, pp. 1408–1436, 2005.

[40] J. L. Taylor, *Several complex variables with connections to algebraic geometry and lie groups*. Providence, RI: AMS, 2002.

Vladislav B. Tadić was born in Belgrade, Serbia, in 1964. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of Belgrade, Belgrade, Serbia, in 1989, 1995, and 1997, respectively.

During 1989–2000, he held various research and teaching positions at the Mihajlo Pupin Institute, Belgrade, Serbia, and the Department of Electrical Engineering, University of Belgrade. From 2000 until 2003, he was a Research Fellow at the Department of Electrical and Electronic Engineering, University of Melbourne, Australia. During 2003–2007, he was a Lecturer at the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, U.K. In 2007, he joined the Department of Mathematics, University of Bristol, Bristol, U.K., where he is currently a Senior Lecturer in Statistics. His research interests include stochastic approximation, dynamic programming, stochastic optimization, and Monte Carlo methods.