The Annals of Statistics
2006, Vol. 34, No. 4, 2026–2068
DOI: 10.1214/009053606000000614

EFFICIENT LIKELIHOOD ESTIMATION IN STATE SPACE MODELS[1]

BY CHENG-DER FUH

National Central University and Academia Sinica

Motivated by studying asymptotic properties of the maximum likelihood estimator (MLE) in stochastic volatility (SV) models, in this paper we investigate likelihood estimation in state space models. We first prove, under some regularity conditions, there is a consistent sequence of roots of the likelihood equation that is asymptotically normal with the inverse of the Fisher information as its variance. With an extra assumption that the likelihood equation has a unique root for each n, then there is a consistent sequence of estimators of the unknown parameters. If, in addition, the supremum of the log likelihood function is integrable, the MLE exists and is strongly consistent. Edgeworth expansion of the approximate solution of likelihood equation is also established. Several examples, including Markov switching models, ARMA models, (G)ARCH models and stochastic volatility (SV) models, are given for illustration.

1. Introduction. Motivated by studying asymptotic properties of the maximum likelihood estimator (MLE) in stochastic volatility (SV) models, in this paper we investigate likelihood estimation in state space models. A state space model is, loosely speaking, a sequence $\{\xi_n\}_{n=0}^{\infty}$ of random variables obtained in the following way. First, a realization of a Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ is created. This chain is sometimes called the regime and is not observed. Then, conditional on $\mathbf{X}$, the ξ-variables are generated. Usually the dependence of ξ_n on $\mathbf{X}$ is more or less local, as when $\xi_n = g(X_n, \xi_{n-1}, \eta_n)$ for some function g and random sequence $\{\eta_n\}$, independent of $\mathbf{X}$. ξ_n itself is generally not Markov and may, in fact, have a complicated dependence structure. When the state space of $\{X_n, n \geq 0\}$ is finite, it is the so-called hidden Markov model or Markov switching model.

The statistical modeling and computation for state space models have attracted a great deal of attention recently because of their importance in applications to speech recognition [49], signal processing [17], ion channels [1], molecular biology [40] and economics [8, 19, 51]. The reader is referred to [20, 34, 41] for a comprehensive summary. The main focus of these efforts has been state space

Received March 2004; revised September 2005.

[1]Supported in part by NSC 95-2118-M-001-008.

AMS 2000 subject classifications. Primary 62M09; secondary 62F12, 62E25.

Key words and phrases. Consistency, efficiency, ARMA models, (G)ARCH models, stochastic volatility models, asymptotic normality, asymptotic expansion, Markov switching models, maximum likelihood, incomplete data, iterated random functions.

modeling and estimation, algorithms for fitting these models and the implementation of likelihood based methods.

The state space model here is defined in a general sense, in which the observations are *conditionally Markovian dependent*, and the state space of the driving Markov chain need not be *finite* or *compact*. When the state space is finite and the observation is a deterministic function of the state space, Baum and Petrie [3] established the consistency and asymptotic normality of the MLE. When the observed random variables are conditionally independent, Leroux [44] proved strong consistency of the MLE, while Bickel, Ritov and Rydén [7] established asymptotic normality of the MLE under mild conditions. Jensen and Petersen [39], Douc and Matias [14] and Douc, Moulines and Rydén [15] studied asymptotic properties of the MLE for general "pseudo-compact" state space models. By extending the inference problem to time series analysis where the state space is finite and the observed random variables are conditionally Markovian dependent, Goldfeld and Quandt [30] and Hamilton [33] considered the implementation of the maximum likelihood estimator in switching autoregressions with Markov regimes. Francq and Roussignol [21] studied the consistency of the MLE, while Fuh [23] established the Bahadur efficiency of the MLE in Markov switching models. We now give two examples of state space models.

EXAMPLE 1 [GARCH(p,q) model]. For given $p \geq 1$ and $q \geq 0$, let

$$Y_n = \sigma_n \varepsilon_n \quad \text{and} \quad \sigma_n^2 = \delta + \sum_{i=1}^{p} \alpha_i \sigma_{n-i}^2 + \sum_{j=1}^{q} \beta_j Y_{n-j}^2, \tag{1.1}$$

where $\delta > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$ are constants, ε_n is a sequence of independent and identically distributed (i.i.d.) random variables, and ε_n is independent of $\{Y_{n-k}, k \geq 1\}$ for all n. This is the celebrated GARCH(p,q) model proposed by Bollerslev [8]. When $q = 0$ or $\beta_j = 0$, for $j = 1, \dots, q$, this is the ARCH(p) model first considered by Engle [19]. The reader is referred to [9] and [20] for a comprehensive summary.

For convenience of notation, we assume that $p, q \geq 2$, and by adding some α_i or β_j equal to zero if necessary. Denote $\eta_n = \sigma_n^{-1} Y_n$, $\tau_n = (\alpha_1 + \beta_1 \eta_n^2, \alpha_2, \dots, \alpha_{p-1}) \in \mathbf{R}^{p-1}$, $\zeta_n = (\eta_n^2, 0, \dots, 0) \in \mathbf{R}^{p-1}$, $\beta = (\beta_2, \dots, \beta_{q-1}) \in \mathbf{R}^{q-2}$, and let I_{p-1} and I_{q-2} be identity matrices. Let A_n be a $(p+q-1) \times (p+q-1)$ matrix written in block form as

$$A_n = \begin{bmatrix} \tau_n & \alpha_p & \beta & \beta_q \\ I_{p-1} & 0 & 0 & 0 \\ \zeta_n & 0 & 0 & 0 \\ 0 & 0 & I_{q-2} & 0 \end{bmatrix}. \tag{1.2}$$

Note that $\{A_n, n \geq 0\}$ are i.i.d. random matrices.

Let $Z = (\delta, 0, \dots, 0)' \in \mathbf{R}^{p+q-1}$ and $X_n = (\sigma_{n+1}^2, \dots, \sigma_{n-p+2}^2, Y_n^2, \dots, Y_{n-q+2}^2)'$, where "$'$" denotes transpose. Following the idea of Bougerol and Picard [10], we have the following state space representation of the GARCH(p, q) model: X_n is a Markov chain governed by

$$X_{n+1} = A_{n+1} X_n + Z, \tag{1.3}$$

and $\xi_n := g(X_n) = (Y_n^2, \dots, Y_{n-q+2}^2)'$, the observed random quantity, is a noninvertible function of X_n.

EXAMPLE 2 (Stochastic volatility models). Let

$$Y_n = \sigma_n \varepsilon_n, \tag{1.4}$$

where $\log \sigma_n^2$ follows an AR(1) process and ε_n is a sequence of i.i.d. random variables with standard normal probability density function. This is the discrete time stochastic volatility model proposed by Taylor [51]. The reader is referred to [29, 50, 52] for a comprehensive summary. Note that Genon-Catalot, Jeantheau and Larédo [27] studied the ergodicity and mixing properties of stochastic volatility models from the hidden Markov model point of view.

Write $X_n := \log \sigma_n^2$ and $Y_n = \sigma \varepsilon_n \exp(X_n/2)$, where σ is a scale parameter. Squaring the observations in the above equation and taking logarithms gives $\log Y_n^2 = \log \sigma^2 + X_n + \log \varepsilon_n^2$. Alternatively, we have

$$\log Y_n^2 = \omega + X_n + \zeta_n, \tag{1.5}$$

where $\omega = \log \sigma^2 + E \log \varepsilon_n^2$, so that the disturbance ζ_n has mean zero by construction. The scale parameter σ also removes the need for a constant term in the stationary first-order autoregressive process

$$X_n = \alpha X_{n-1} + \eta_n, \qquad |\alpha| < 1, \tag{1.6}$$

where η_n is a sequence of i.i.d. random variables distributed as $N(0, \sigma_\eta^2)$. Moreover, we assume that ζ_n and η_n are independent. Note that in (1.5) and (1.6) the observed random quantity is $\xi_n := \log Y_n^2$. $\{X_n, n \geq 0\}$ and forms a Markov chain with transition probability

$$p(x_{k-1}, x_k) = (2\pi\sigma_\eta^2)^{-1/2} \exp\left\{-\frac{1}{2} \frac{(x_k - \alpha x_{k-1})^2}{\sigma_\eta^2}\right\} \tag{1.7}$$

and stationary distribution $\pi \sim N(0, \sigma_\eta^2/(1-\alpha))$.

For given observations $\mathbf{y} = (\log y_1^2, \dots, \log y_n^2)$ from the state space model (1.5) and (1.6), the likelihood function of the parameter $\theta = (\alpha, \sigma_\eta^2)$ is

$$\begin{aligned} l(\mathbf{y}; \theta) = \int_{x_0 \in \mathcal{X}} \cdots \int_{x_n \in \mathcal{X}} \pi(x_0) \prod_{k=1}^n p(x_{k-1}, x_k) \\ \times f_\zeta(\log y_k^2 - \omega - x_k)\, dx_n \cdots dx_0, \end{aligned} \tag{1.8}$$

where $f_\zeta(\cdot)$ is the probability density function of ζ_1.

A major difficulty in analyzing the likelihood function in state space models is that it can be expressed only in integral form; see equation (1.8), for instance. In this paper we provide a device which represents the integral likelihood function as the L_1-norm of a Markovian iterated random functions system. This new representation enables us to apply results of the strong law of large numbers, central limit theorem and Edgeworth expansion for the distributions of Markov random walks, and to verify strong consistency of the MLE and first-order efficiency and Edgeworth expansion on the solution of the likelihood equation. Note that third-order efficiency follows from Edgeworth expansion by a standard argument (cf. [28]). Another essential point worth being mentioned is that we introduce *a weight function* in a suitable way [see (4.1)–(4.3), Assumptions K2, K3 and Definition 2 in Section 4, and C1 in Section 5] to relax the condition of a compact state space for the underlying Markov chain, and to cover several interesting examples.

The remainder of this paper is organized as follows. In Section 2 we define the state space model as a general state Markov chain in a Markovian random environment, and represent the likelihood function as the L_1-norm of a Markovian iterated random functions system. In Section 3 we give a brief summary of a Markovian iterated random functions system, and provide an ergodic theorem and the strong law of large numbers. The multivariate central limit theorem and Edgeworth expansion for a Markovian iterated random functions system are given in Section 4. Section 5 contains our main results, where we consider efficient likelihood estimation in state space models, and state the main results. First, we compute Fisher information and prove the existence of an efficient estimator in a "Cramér fashion." Second, we characterize Kullback–Leibler information, and prove strong consistency of the MLE. Last, we establish Edgeworth expansion of the approximate solution of the likelihood equation. In Section 6 we consider a few examples, including Markov switching models, ARMA models, (G)ARCH models and SV models, which are commonly used in financial economics. The proofs of the lemmas in Section 5 are given in Section 7. Other technical proofs are deferred to the Appendix.

2. State space models. A state space model is defined as a parameterized Markov chain in a Markovian random environment with the underlying environmental Markov chain viewed as missing data. Specifically, let $\mathbf{X} = \{X_n, n \geq 0\}$ be a Markov chain on a general state space $\mathcal{X}$, with transition probability kernel $P^\theta(x, \cdot) = P^\theta\{X_1 \in \cdot | X_0 = x\}$ and stationary probability $\pi_\theta(\cdot)$, where $\theta \in \Theta \subseteq \mathbf{R}^q$ denotes the unknown parameter. Suppose that a random sequence $\{\xi_n\}_{n=0}^\infty$, taking values in $\mathbf{R}^d$, is adjoined to the chain such that $\{(X_n, \xi_n), n \geq 0\}$ is a Markov chain on $\mathcal{X} \times \mathbf{R}^d$ satisfying $P^\theta\{X_1 \in A | X_0 = x, \xi_0 = s\} = P^\theta\{X_1 \in A | X_0 = x\}$ for $A \in \mathcal{B}(\mathcal{X})$, the σ-algebra of $\mathcal{X}$. And conditioning on the full $\mathbf{X}$ sequence, ξ_n is a Markov chain with probability

$$P^{\theta}\{\xi_{n+1} \in B | X_0, X_1, \ldots; \xi_0, \xi_1, \ldots, \xi_n\} \tag{2.1}$$
$$= P^{\theta}\{\xi_{n+1} \in B | X_{n+1}; \xi_n\} \qquad \text{a.s.}$$

for each n and $B \in \mathcal{B}(\mathbf{R}^d)$, the Borel σ-algebra on $\mathbf{R}^d$. Note that in (2.1) the conditional probability of ξ_{n+1} depends on X_{n+1} and ξ_n only. Furthermore, we assume the existence of a transition probability density $p_\theta(x, y)$ for the Markov chain $\{X_n, n \geq 0\}$ with respect to a σ-finite measure m on $\mathcal{X}$ such that

$$P^{\theta}\{X_1 \in A, \xi_1 \in B | X_0 = x, \xi_0 = s_0\} \tag{2.2}$$
$$= \int_{y \in A} \int_{s \in B} p_\theta(x, y) f(s; \theta | y, s_0) Q(ds) m(dy),$$

where $f(\xi_k; \theta | X_k, \xi_{k-1})$ is the conditional probability density of ξ_k given ξ_{k-1} and X_k, with respect to a σ-finite measure Q on $\mathbf{R}^d$. We also assume that the Markov chain $\{(X_n, \xi_n), n \geq 0\}$ has a stationary probability with probability density function $\pi(x) f(\cdot; \theta | x)$ with respect to $m \times Q$. In this paper we consider $\theta = (\theta_1, \ldots, \theta_q) \in \Theta \subseteq \mathbf{R}^q$ as the unknown parameter, and the true parameter value is denoted by θ_0. We will use $\pi(x)$ for $\pi_\theta(x)$, $p(x, y)$ for $p_\theta(x, y)$, $f(\xi_0 | X_0)$ for $f(\xi_0; \theta | X_0)$, and $f(\xi_k | X_k, \xi_{k-1})$ for $f(\xi_k; \theta | X_k, \xi_{k-1})$, here and in the sequel, depending on our convenience. Now we give a formal definition as follows.

DEFINITION 1. $\{\xi_n, n \geq 0\}$ is called a state space model if there is a Markov chain $\{X_n, n \geq 0\}$ such that the process $\{(X_n, \xi_n), n \geq 0\}$ satisfies (2.1).

Note that this setting includes several interesting examples of Markov-switching Gaussian autoregression of Hamilton [33], (G)ARCH models of Engle [19] and Bollerslev [8], and SV models of Clark [12] and Taylor [51]. When the state space $\mathcal{X}$ is finite or compact, this reduces to the hidden Markov model considered by Francq and Roussignol [21], Fuh [22, 23, 25] and Douc, Moulines and Rydén [15]. Denote $S_n = \sum_{t=1}^n \xi_t$. When ξ_n are *conditionally independent* given $\mathbf{X}$, the Markov chain $\{(X_n, S_n), n \geq 0\}$ is called a *Markov additive process* and S_n is called a *Markov random walk*. Furthermore, if the state space $\mathcal{X}$ is finite, $\{\xi_n, n \geq 0\}$ is the hidden Markov model studied by Leroux [44], Bickel and Ritov [6] and Bickel, Ritov and Rydén [7]. When the state space $\mathcal{X}$ is "pseudo-compact" and ξ_n are conditionally independent given $\mathbf{X}$, $\{\xi_n, n \geq 0\}$ is the state space model considered in [39] and [14].

For given observations $s_0, s_1, \ldots, s_n$ from a state space model $\{\xi_n, n \geq 0\}$, the likelihood function is

$$
\begin{aligned}
& p_n(s_0, s_1, \ldots, s_n; \theta) \\
& \quad = \int_{x_0 \in \mathcal{X}} \cdots \int_{x_n \in \mathcal{X}} \pi_\theta(x_0) f(s_0; \theta | x_0) \\
& \qquad \times \prod_{j=1}^{n} p_\theta(x_{j-1}, x_j) \\
& \qquad \times f(s_j; \theta | x_j, s_{j-1}) m(dx_n) \cdots m(dx_0).
\end{aligned} \tag{2.3}
$$

Recall that $\pi_\theta(x_0) f(s_0; \theta | x_0)$ is the stationary probability density with respect to $m \times Q$ of the Markov chain $\{(X_n, \xi_n), n \geq 0\}$.

To represent the likelihood $p_n(\xi_0, \xi_1, \ldots, \xi_n; \theta)$ as the L_1-norm of a Markovian iterated random functions system, let

$$
\mathbf{M} = \left\{ h | h : \mathcal{X} \to \mathbf{R}^+ \text{ is } m\text{-measurable and } \int_{x \in \mathcal{X}} h(x) m(dx) < \infty \right\}. \tag{2.4}
$$

For each $j = 1, \ldots, n$, define the random functions $\mathbf{P}_\theta(\xi_0)$ and $\mathbf{P}_\theta(\xi_j)$ on $(\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$ as

$$
\mathbf{P}_\theta(\xi_0) h(x) = \int_{x \in \mathcal{X}} f(\xi_0; \theta | x) h(x) m(dx), \qquad \text{a constant,} \tag{2.5}
$$

$$
\mathbf{P}_\theta(\xi_j) h(x) = \int_{y \in \mathcal{X}} p_\theta(x, y) f(\xi_j; \theta | y, \xi_{j-1}) h(y) m(dy). \tag{2.6}
$$

Define the composition of two random functions as

$$
\begin{aligned}
& \mathbf{P}_\theta(\xi_{j+1}) \circ \mathbf{P}_\theta(\xi_j) h(x) \\
& \quad = \int_{z \in \mathcal{X}} p_\theta(x, z) f(\xi_j; \theta | z, \xi_{j-1}) \\
& \qquad \times \left(\int_{y \in \mathcal{X}} p_\theta(z, y) f(\xi_{j+1}; \theta | y, \xi_j) h(y) m(dy) \right) m(dz).
\end{aligned} \tag{2.7}
$$

For $h \in \mathbf{M}$, denote $\|h\| := \int_{x \in \mathcal{X}} h(x) m(dx)$ as the L^1-norm on $\mathbf{M}$ with respect to m. Then the likelihood $p_n(\xi_0, \xi_1, \ldots, \xi_n; \theta)$ can be represented as

$$
\begin{aligned}
& p_n(\xi_0, \xi_1, \ldots, \xi_n; \theta) \\
& \quad = \int_{x_0 \in \mathcal{X}} \cdots \int_{x_n \in \mathcal{X}} \pi_\theta(x_0) f(\xi_0; \theta | x_0) \\
& \qquad \times \prod_{j=1}^{n} p_\theta(x_{j-1}, x_j) \\
& \qquad \times f(\xi_j; \theta | x_j, \xi_{j-1}) m(dx_n) \cdots m(dx_0) \\
& \quad = \|\mathbf{P}_\theta(\xi_n) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0) \pi_\theta\|.
\end{aligned} \tag{2.8}
$$

Note that, for $j = 1, \dots, n$, the integrand $p_\theta(x, y) f(\xi_j; \theta | y, \xi_{j-1})$ of $\mathbf{P}_\theta(\xi_j)$ in (2.6) and (2.8) represents $X_{j-1} = x$ and $X_j \in dy$, and ξ_j is a Markov chain with transition probability density $f(\xi_j; \theta | y, \xi_{j-1})$ for given $\mathbf{X}$. By definition (2.1), $\{(X_n, \xi_n), n \geq 0\}$ is a Markov chain, and this implies that $\mathbf{P}_\theta(\xi_j)$ is a sequence of Markovian iterated random functions systems (see Section 5 for a formal definition). Therefore, by representation (2.8), $p_n(\xi_0, \xi_1, \dots, \xi_n; \theta)$ is the L_1-norm of a Markovian iterated random functions system.

3. Ergodic theorems for a Markovian iterated random functions system. To analyze the asymptotic properties of efficient likelihood estimators in state space models, in this section we study the ergodic theorem and the strong law of large numbers for a Markovian iterated random functions system. The Markovian iterated random functions system is a generalization of an iterated random functions system, in which the random functions are driven by a Markov chain. For a general account of an iterated random functions system, the reader is referred to [13] for a recent survey.

For simplicity in our notation, let $\{Y_n, n \geq 0\}$ [instead of $\{(X_n, \xi_n), n \geq 0\}$ in Section 2] be a Markov chain on a general state space $\mathcal{Y}$ with σ-algebra $\mathcal{A}$, which irreducible with respect to a maximal irreducibility measure on $(\mathcal{Y}, \mathcal{A})$ and is aperiodic. The transition kernel is denoted by $P(y, A)$. Let $(\mathbf{M}, d)$ be a complete separable metric space with Borel σ-algebra $\mathcal{B}(\mathbf{M})$. Denote by M_0 a random variable which is independent of $\{Y_n, n \geq 0\}$. A sequence of the form

$$M_n = F(Y_n, M_{n-1}), \qquad n \geq 1, \tag{3.1}$$

taking values in $(\mathbf{M}, d)$ is called a *Markovian iterated random functions system* (MIRFS) of Lipschitz functions providing the following:

(1) $\{Y_n, n \geq 0\}$ is a Markov chain taking values in a second countable measurable space $(\mathcal{Y}, \mathcal{A})$, with transition probability kernel $P(\cdot, \cdot)$ and stationary probability π, and M_0 is a random element on a probability space $(\Omega, \mathcal{F}, P)$, which is independent of $\{Y_n, n \geq 0\}$;

(2) $F : (\mathcal{Y} \times \mathbf{M}, \mathcal{A} \otimes \mathcal{B}(\mathbf{M})) \to (\mathbf{M}, \mathcal{B}(\mathbf{M}))$ is jointly measurable and Lipschitz continuous in the second argument.

Clearly, $\{(Y_n, M_n), n \geq 0\}$ constitutes a Markov chain with state space $\mathcal{Y} \times \mathbf{M}$ and transition probability kernel $\mathbf{P}$, given by

$$\mathbf{P}\big((y, u), A \times B\big) := \int_{z \in A} I_B\big(F(z, u)\big) P(y, dz) \tag{3.2}$$

for all $y \in \mathcal{Y}, u \in \mathbf{M}, A \in \mathcal{A}$ and $B \in \mathcal{B}(\mathbf{M})$, where I denotes the indicator function. The n-step transition kernel is denoted $\mathbf{P}^n$. For $(y, u) \in \mathcal{Y} \times \mathbf{M}$, let $\mathbf{P}_{yu}$ be the probability measure on the underlying measurable space under which $Y_0 = y, M_0 = u$ a.s. The associated expectation is denoted $\mathbf{E}_{yu}$, as usual. For an arbitrary distribution ν on $\mathcal{Y} \times \mathbf{M}$, we put $\mathbf{P}_\nu(\cdot) := \int \mathbf{P}_{yu}(\cdot)\nu(dy \times du)$ with associated expectation $\mathbf{E}_\nu$. We use $\mathbf{P}$ and $\mathbf{E}$ for probabilities and expectations, respectively, that do not depend on the initial distribution.

Let $\mathbf{M}_0$ be a dense subset of $\mathbf{M}$ and $\mathcal{M}(\mathbf{M_0}, \mathbf{M})$ the space of all mappings $h : \mathbf{M_0} \to \mathbf{M}$ endowed with the product topology and product σ-algebra. Then the space $\mathcal{L}_{\mathrm{Lip}}(\mathbf{M}, \mathbf{M})$ of all Lipschitz continuous mappings $h : \mathbf{M} \to \mathbf{M}$ properly embedded forms a Borel subset of $\mathcal{M}(\mathbf{M}_0, \mathbf{M})$, and the mappings

$$\mathcal{L}_{\mathrm{Lip}}(\mathbf{M}, \mathbf{M}) \times \mathbf{M} \ni (h, u) \mapsto h(u) \in \mathbf{M},$$

$$\mathcal{L}_{\mathrm{Lip}}(\mathbf{M}, \mathbf{M}) \ni h \mapsto l(h) := \sup_{u \neq v} \frac{d(h(u), h(v))}{d(u, v)}$$

are Borel; see Lemma 5.1 in [13] for details. Hence,

$$L_n := l\big(F(Y_n, \cdot)\big), \qquad n \geq 0, \tag{3.3}$$

are also measurable and form a sequence of Markovian dependent random variables.

An important point to characterize the limit in the ergodic theorem will be the right use of the idea of duality. For this purpose, we introduce a time-reversed (or dual) Markov chain $\{\tilde{Y}_n, n \geq 0\}$ of $\{Y_n, n \geq 0\}$ as follows. Assume that there exists a σ-finite measure m on $(\mathcal{Y}, \mathcal{A})$ such that the probability measure P on $(\mathcal{Y}, \mathcal{A})$ defined by $P(A) = P(Y_1 \in A | Y_0 = y)$ is absolutely continuous with respect to m, so that $P(A) = \int_A p(y, z) m(dz)$ for all $A \in \mathcal{A}$, where $p(y, \cdot) = dP/dm$. The Markov chain $\{Y_n, n \geq 0\}$ is assumed to have an invariant probability measure π which has a positive probability density function π (without any confusion, we still use the same notation) with respect to m. We shall use $\sim$ to refer to the time-reversed (or dual) process $\{\tilde{Y}_n, n \geq 0\}$ with transition probability density

$$\tilde{p}(z, y) = p(y, z)\pi(y)/\pi(z). \tag{3.4}$$

Denote $\tilde{P}$ as the corresponding probability. It is easy to see that both Y_n and $\tilde{Y}_n$ have the same stationary distribution π. In this section we will assume that the initial distribution of Y_0 is the stationary distribution π.

In the following, we write $F_n(u)$ for $F(Y_n, u)$. For all $1 \leq k \leq n$, let $F_{k:n} := F_k \circ \cdots \circ F_n$, $F_{n:k} := F_n \circ \cdots \circ F_k$, where $\circ$ denotes the composition of functions. Denote $F_{n:n-1}$ as the identity on $\mathbf{M}$, Hence

$$M_n = F_n(M_{n-1}) = F_{n:1}(M_0) \tag{3.5}$$

for all $n \geq 0$. Closely related to these *forward iterations*, and in fact a key tool to the analysis of the ergodic property, is the sequence of *backward iterations*

$$\tilde{M}_n := F_{1:n}(M_0), \qquad n \geq 0. \tag{3.6}$$

The connection is established by the identity

$$\pi(y)\mathbf{P}(M_n \in \cdot | Y_0 = y) = \pi(z)\tilde{\mathbf{P}}(\tilde{M}_n \in \cdot | \tilde{Y}_0 = z) \tag{3.7}$$

for all $n \geq 0$. Put also $M_n^u := F_{n:1}(u)$ and $\tilde{M}_n^u := F_{1:n}(u)$ for $u \in \mathbf{M}$ and note that

$$\begin{aligned}
&\int_{z\in\mathcal{Y}}\int_{y\in\mathcal{Y}} \mathbf{P}\big((M_n^u, \tilde{M}_n^u)_{n\geq 0} \in \cdot | Y_0 = y, \tilde{Y}_0 = z\big)\pi(dy)\pi(dz) \\
&\quad = \int_{z\in\mathcal{Y}}\int_{y\in\mathcal{Y}} \mathbf{P}\big((M_n, \tilde{M}_n)_{n\geq 0} \in \cdot | Y_0 = y, \tilde{Y}_0 = z\big)\pi(dy)\pi(dz).
\end{aligned} \tag{3.8}$$

Note that in (3.8), the probability $\mathbf{P}$ denotes a joint probability.

$\{Y_n, n \geq 0\}$ is called *Harris recurrent* if there exist a set $A \in \mathcal{A}$, a probability measure Γ concentrated on A and an ε with $0 < \varepsilon < 1$ such that $P_y(Y_n \in A \text{ i.o.}) = 1$ for all $y \in \mathcal{Y}$ and, furthermore, there exists n such that $P^n(y, A') \geq \varepsilon\Gamma(A')$ for all $y \in A$ and all $A' \in \mathcal{A}$.

A central question for an MIRFS $(M_n)_{n\geq 0}$ is under which conditions it stabilizes, that is, converges to a stationary distribution Π. The next theorem summarizes the results regarding this question.

THEOREM 1. *Let $\{Y_n, n \geq 0\}$ be an aperiodic, irreducible and Harris recurrent Markov chain, and let $(M_n)_{n\geq 0}$ be an MIRFS of Lipschitz functions. Suppose the initial distribution of Y_0 is π, and*

$$\mathbf{E}\log l(F_1) < 0 \quad \textit{and} \quad \mathbf{E}\log^+ d\big(F_1(u_0), u_0\big) < \infty \tag{3.9}$$

for some $u_0 \in \mathbf{M}$. Then the following assertions hold:

(i) *$\tilde{M}_n$ converges a.s. to a random element $\tilde{M}_\infty$ which does not depend on the initial distribution.*

(ii) *M_n converges in distribution to $\tilde{M}_\infty$ under $\mathbf{P}$.*

(iii) *Define Π as the stationary distribution of $(\tilde{Y}_\infty, \tilde{M}_\infty)$. Then Π is the unique stationary probability of the Markov chain $\{(Y_n, M_n), n \geq 0\}$.*

(iv) *$(M_n)_{n\geq 0}$ is ergodic under $\mathbf{P}_\Pi$, that is, for any $u \in \mathbf{M}$,*

$$\frac{1}{n}\sum_{k=1}^{n} g(M_k) \longrightarrow \mathbf{E}_\Pi(g(\tilde{M}_\infty)), \qquad \mathbf{P}_\Pi\text{-}a.s. \tag{3.10}$$

for all bounded continuous real-valued functions g on $\mathbf{M}$.

We remark that Elton [18] showed in the situation of a stationary sequence $(F_n)_{n\geq 1}$ that Theorem 1 holds whenever $\mathbf{E}\log^+ l(F_1)$ and $\mathbf{E}\log^+ d(F_1(u_0), u_0)$ are both finite for some (and then all) $u_0 \in \mathbf{M}$ and the Lyapunov exponent $\gamma := \lim_{n\to\infty} n^{-1}\log l(F_{n:1})$, which exists by Kingman's subadditive ergodic theorem, is a.s. negative. Since the initial distribution of Y_0 is the stationary distribution π, the Markov chain Y_n is a stationary sequence, and hence, M_n is a sequence of iterated random functions generated by stationary sequences. Here, we impose the Harris recurrent condition so that the invariant measure π exists, and we are able to characterize $\tilde{M}_\infty$ in a Markovian setting. Since the proof is similar to that in [2], it is omitted.

4. Central limit theorem and Edgeworth expansion for distributions of a Markovian iterated random functions system. Consider the Markovian iterated random functions system $\{(Y_n, M_n), n \geq 0\}$ defined in (3.1). Abuse the notation a little bit and let g be an $\mathbf{R}^p$-valued function on $\mathbf{M}$. In this section we study the central limit theorem and Edgeworth expansion of the sum $S_n = \sum_{k=1}^n g(M_k)$ and $\mathbf{g}(n^{-1}S_n)$ for a smooth function $\mathbf{g}: \mathbf{R}^p \to \mathbf{R}^q$. Let $w: \mathcal{Y} \to [1, \infty)$ be a measurable function, and let $\mathbf{B}$ be the Banach space of measurable functions $h: \mathcal{Y} \to C$ (:= the set of complex numbers) with $\|h\|_w := \sup_y |h(y)|/w(y) < \infty$. Assume further that $\{Y_n, n \geq 0\}$ has a stationary distribution π with $\int w(y)\pi(dy) < \infty$, and

$$\lim_{n\to\infty} \sup_y \left\{ \left| E[h(Y_n)|Y_0 = y] - \int h(z)\pi(dz) \right| \Big/ w(y) : y \in \mathcal{Y}, |h| \leq w \right\} = 0, \tag{4.1}$$

$$\sup_y \{E[w(Y_p)|Y_0 = y]/w(y)\} < \infty, \tag{4.2}$$

for some $p \geq 1$. Condition (4.1) says that the chain is w-uniformly ergodic, which implies that there exist $\gamma > 0$ and $0 < \rho < 1$ such that, for all $h \in \mathbf{B}$ and $n \geq 1$,

$$\sup_y \left| E[h(Y_n)|Y_0 = y] - \int h(z)\pi(dz) \right| \Big/ w(y) \leq \gamma\rho^n \|h\|_w, \tag{4.3}$$

(cf. pages 382–383 and Theorem 16.0.1 of [46]). We remark that, for $w = 1$, condition (4.1) is the classical uniform ergodicity condition for $\{Y_n, n \geq 0\}$.

The following assumption will be assumed throughout this section.

ASSUMPTION K.

K1. Let $\{Y_n, n \geq 0\}$ be an aperiodic, irreducible Markov chain satisfying conditions (4.1)–(4.2). Furthermore, we assume the initial distribution of Y_0 is π.

K2. The MIRFS $(M_n)_{n\geq 0}$ has the weighted mean contraction property, that is, there exists a $p \geq 1$ such that

$$\sup_y \left\{ \mathbf{E}\left(\log \frac{L_p w(Y_p)}{w(y)} \Big| Y_0 = y \right) \right\} < 0.$$

K3. There exists $u_0 \in \mathbf{M}$ for which

$$\mathbf{E}d^2(F_1(u_0), u_0) < \infty \quad \text{and} \quad \sup_y \left\{ \mathbf{E}\left(\frac{L_1 w(Y_1)}{w(y)} \Big| Y_0 = y \right) \right\} < \infty.$$

REMARK 1. (a) Assumption K1 is a condition for the underlying Markov chain $\{Y_n, n \geq 0\}$ which is general enough to include several practical used models studied in Section 6. Assumption K2 is a weighted mean contraction condition which is different from the standard mean contraction condition $\mathbf{E}\log L_1 < 0$ used in Theorem 1. Assumption K3 is a weighted moment condition. Note that under Assumptions K1–K3, and the extra assumption that $\{(Y_n, M_n),$

$n \geq 0\}$ is an irreducible, aperiodic and Harris recurrent Markov chain, Theorems 13.0.1 and 17.0.1(i) of [46] imply that Theorem 1 still holds. Furthermore, we will prove the central limit theorem and Edgeworth expansion for the distributions of a Markovian iterated random functions system in Theorem 2.

(b) To have better understanding of Assumption K, we consider a simple state space model. Given $p \geq 1$ as in Assumption K2, and $|\alpha| < 1$, let $Y_n = \alpha Y_{n-1} + \varepsilon_n$, $\xi_n = \beta_{Y_n}\xi_{n-1} + \eta_n$, where ε_n are i.i.d. random variables with $E|\varepsilon_1| = c < \infty$, and η_n are i.i.d. random variables with $E|\eta_1| < \infty$. Further, we assume both ε_1 and η_1 have positive probability density function with respect to Lebesgue measure, and that they are mutually independent. Denote $b = (1 - |\alpha|^p)/(1 - |\alpha|)$ and $a = 1/(bc + 1) < 1$, and assume $|\beta_y| < a^{1/p} < 1$ for all $y \in \mathscr{Y}$. It is known that $w(y) = |y| + 1$ (cf. pages 380 and 383 of [46]). Let $d(u, v) = |u - v|$. It is easy to see that Assumption K1 and the first part of Assumption K3 hold. To check Assumption K2, we have

$$\begin{aligned}
&\sup_y \left\{ \mathbf{E}\left(\log \frac{L_p w(Y_p)}{w(y)} \Big| Y_0 = y \right) \right\} \\
&\quad = \sup_y \left\{ \mathbf{E}\left(\log \frac{|\beta_{Y_p} \cdots \beta_{Y_1}|(|\alpha^p y + \sum_{k=0}^{p-1} \alpha^k \varepsilon_{p-k}| + 1)}{|y| + 1} \Big| Y_0 = y \right) \right\} \\
&\quad < \log \sup_y \left\{ \frac{a(|\alpha^p y| + E|\sum_{k=0}^{p-1} \alpha^k \varepsilon_{p-k}| + 1)}{|y| + 1} \right\} \\
&\quad = \log \sup_y \left\{ \frac{a(|\alpha^p y| + bc + 1)}{|y| + 1} \right\} = 0.
\end{aligned} \tag{4.4}$$

By using the same argument, we have the second part of Assumption K3. When ε_n are i.i.d. $N(0, 1)$, η_n are i.i.d. $N(0, 1)$, and they are mutually independent. Then $a = \sqrt{2\pi}/(2b + \sqrt{2\pi}\,) < 1$.

Recall that Π is defined in Theorem 1(iii) and denote $Q(B) := \Pi(\mathscr{Y} \times B)$ for all $B \in \mathscr{B}(\mathbf{M})$. Let $g \in \mathscr{L}_0^2(Q)$ be a square integrable function taking values in $\mathbf{R}^p$ with mean $\mathbf{0}$, that is, $g = (g_1, \ldots, g_p)$ with each g_k a real-valued function on $\mathbf{M}$, and

$$\int_{\mathbf{M}} g_k(u) Q(du) = 0, \qquad \|g_k\|_2^2 = \int_{\mathbf{M}} g_k^2(u) Q(du) < \infty, \tag{4.5}$$

for $k = 1, \ldots, p$. Consider the sequence

$$S_n = S_n(g) = g(M_1) + \cdots + g(M_n), \qquad n \geq 1, \tag{4.6}$$

which may be viewed as a Markov random walk on the Markov chain $\{(Y_n, M_n), n \geq 0\}$.

Note that there are two special properties of the Markov chain induced by the Markovian iterated random functions system (2.4)–(2.7). First, the hypothesis that the transition probability possesses a density leads to a classical situation in the context of the so-called "Doeblin condition" for Markov chains. Second, a positivity hypothesis on $\mathbf{M}$ defined in (2.4) in the support of the Markov chain leads to contraction properties, on which basis we will develop the spectral theory. The reader is referred to [37] for a general account of the perturbation theory of Markovian operators. We need the following notation first.

DEFINITION 2. Let $w : \mathcal{Y} \to [1, \infty)$ be a weight function. For any measurable function $\varphi : \mathcal{Y} \times \mathbf{M} \to [1, \infty)$, given $u_0 \in \mathbf{M}$, define

$$\|\varphi\|_w := \sup_{y \in \mathcal{Y}, u \in \mathbf{M}} \frac{|\varphi(y, u)|}{w(y)}$$

and

$$\|\varphi\|_h := \sup_{y \in \mathcal{Y}, u, v : 0 < d(u,v) \le 1} \frac{|\varphi(y, u) - \varphi(y, v)|}{(w(y)\, d(u, v))^\delta},$$

for $0 < \delta < 1$. We define $\mathcal{H}$ as the set of φ on $\mathcal{Y} \times \mathbf{M}$ for which $\|\varphi\|_{wh} := \|\varphi\|_w + \|\varphi\|_h$ is finite, where wh represents a combination of the weighted variation norm and the bounded weighted Hölder norm.

Let ν be an initial distribution of (Y_0, M_0) and let $\mathbf{E}_\nu$ denote expectation under the initial distribution ν on (Y_0, M_0). For $\varphi \in \mathcal{H}$, $g \in \mathcal{L}^2(Q)$, $y \in \mathcal{Y}$, $u \in \mathbf{M}$ and $p \times 1$ vectors $\alpha = (\alpha_1, \dots, \alpha_p)' \in \mathbf{R}^p$, define linear operators $\mathbf{T}_\alpha$, $\mathbf{T}$, ν_α and $\mathbf{Q}$ on the space $\mathcal{H}$ as

$$(\mathbf{T}_\alpha \varphi)(y, u) = \mathbf{E}\{e^{i\alpha' g(M_1)} \varphi(Y_1, M_1) | Y_0 = y, M_0 = u\}, \tag{4.7}$$

$$(\mathbf{T}\varphi)(y, u) = \mathbf{E}\{\varphi(Y_1, M_1) | Y_0 = y, M_0 = u\}, \tag{4.8}$$

$$\nu_\alpha \varphi = \mathbf{E}_\nu\{e^{i\alpha' \varphi(u)} \varphi(Y_0, u)\}, \qquad \mathbf{Q}\varphi = \mathbf{E}_\Pi\{\varphi(Y_0, u)\}. \tag{4.9}$$

In the case of a w-uniformly ergodic Markov chain, Fuh and Lai [26] have shown that there exists a sufficiently small $\delta > 0$ such that, for $|\alpha| \le \delta$, $\mathcal{H} = \mathcal{H}_1(\alpha) \oplus \mathcal{H}_2(\alpha)$ and

$$\mathbf{T}_\alpha \mathbf{Q}_\alpha \varphi = \lambda(\alpha) \mathbf{Q}_\alpha \varphi \qquad \text{for all } \varphi \in \mathcal{H}, \tag{4.10}$$

where $\mathcal{H}_1(\alpha)$ is a one-dimensional subspace of $\mathcal{H}$, $\lambda(\alpha)$ is the eigenvalue of $\mathbf{T}_\alpha$ with corresponding eigenspace $\mathcal{H}_1(\alpha)$ and $\mathbf{Q}_\alpha$ is the parallel projection of $\mathcal{H}$ onto the subspace $\mathcal{H}_1(\alpha)$ in the direction of $\mathcal{H}_2(\alpha)$. Extension of their argument to the weight functions w and l defined in Definition 2 is given in the Appendix, which also proves the following lemmas.

LEMMA 1. *Let* $\{(Y_n, M_n), n \ge 0\}$ *be the MIRFS of Lipschitz functions defined in* (2.1) *and satisfying Assumption* K. *Assume* $g \in \mathcal{L}^r(Q)$ *for some* $r > 2$.

Then $\mathbf{T}$ *and* $\mathbf{Q}$ *are bounded linear operators on the Banach space* $\mathcal{H}$ *with norm* $\|\cdot\|_{wh}$, *and satisfy*

$$\|\mathbf{T}^n - \mathbf{Q}\|_{wh} = \sup_{\varphi \in \mathcal{H}, \|\varphi\|_{wh} \le 1} \|\mathbf{T}^n \varphi - \mathbf{Q}\varphi\|_{wh} < \gamma_* \rho_*^n, \tag{4.11}$$

for some $\gamma_* > 0$ *and* $0 < \rho_* < 1$.

By using an argument similar to Proposition 1 of [24], we have the following:

LEMMA 2. *Let* $\{(Y_n, M_n), n \ge 0\}$ *be the MIRFS defined in* (2.1) *satisfying Assumption* K, *such that the induced Markov chain* $\{(Y_n, M_n), n \ge 0\}$ *with transition probability kernel* (3.2) *is irreducible*, *aperiodic and Harris recurrent. Assume* $g \in \mathcal{L}^r(Q)$ *for some* $r > 2$. *Then there exists* $\delta > 0$ *such that*, *for* $\alpha \in \mathbf{R}^p$ *with* $|\alpha| < \delta$, *and for* $\varphi \in \mathcal{H}$,

$$\begin{aligned} \mathbf{E}_\nu\{e^{i\alpha' g(M_n)}\varphi(Y_n, M_n)\} &= \nu_\alpha \mathbf{T}_\alpha^n \varphi = \nu_\alpha \mathbf{T}_\alpha^n \{\mathbf{Q}_\alpha + (I - \mathbf{Q}_\alpha)\}\varphi \\ &= \lambda^n(\alpha)\nu_\alpha \mathbf{Q}_\alpha \varphi + \nu_\alpha \mathbf{Q}_\alpha^n (I - \mathbf{Q}_\alpha)\varphi, \end{aligned} \tag{4.12}$$

and:

(i) $\lambda(\alpha)$ *is the unique eigenvalue of the maximal modulus of* $\mathbf{T}_\alpha$;
(ii) $\mathbf{Q}_\alpha$ *is a rank-one projection*;
(iii) *the mappings* $\lambda(\alpha)$, $\mathbf{Q}_\alpha$ *and* $I - \mathbf{Q}_\alpha$ *are analytic*;
(iv) $|\lambda(\alpha)| > \frac{2+\rho_*}{3}$ *and for each* $k \in N$, *the set of positive integers*, *there exists* $c > 0$ *such that*, *for each* $n \in N$ *and* $j_1, \dots, j_p$ *with* $j_1 + \cdots + j_p = k$,

$$\left\| \frac{\partial^k}{\partial \alpha_1^{j_1} \cdots \partial \alpha_p^{j_p}} (I - \mathbf{Q}_\alpha)^n \right\|_{wh} \le c \left(\frac{1 + 2\rho_*}{3} \right)^n;$$

(v) *denote* $g = (g_1, \dots, g_p)$, *and let* $\gamma_j := \lim_{n\to\infty}(1/n)\mathbf{E}_{yu} \log \|g_j(M_n)\|$, *the upper Lyapunov exponent*; *it follows that*

$$\gamma_j = \frac{\partial \lambda(\alpha)}{\partial \alpha_j}\bigg|_{\alpha=0} = \int \mathbf{E}_{yu} g_j(M_1)\Pi(dy \times du). \tag{4.13}$$

Note that in Lemma 2 we need the extra assumption that the induced Markov chain $\{(Y_n, M_n), n \ge 0\}$ with transition probability kernel (3.2) is irreducible, aperiodic and Harris recurrent. In Section 5 we will show that this condition is satisfied for the Markov chain induced by the Markovian iterated random functions system (2.4)–(2.7).

For given $S_n = \sum_{k=1}^n g(M_k)$ of the MIRFS $\{(Y_n, M_n), n \ge 0\}$, in this section we will obtain Edgeworth expansions for the standardized distribution of S_n via the representation (4.12) of the characteristic function $\mathbf{E}(e^{i\alpha' g(M_n)} | Y_0 = y, M_0 = 0)$. Note that Lemma 1 implies that $\{(Y_n, M_n), n \ge 0\}$ is geometrically mixing in the sense that there exist $r_1 > 0$ and $0 < \gamma_1 < 1$ such that, for all $y \in \mathcal{Y}$, $u \in \mathbf{M}$, $k \ge 0$

and $n \geq 1$ and for all real-valued measurable functions φ_1, φ_2 with $\|\varphi_1^2\|_{wh} < \infty$ and $\|\varphi_2^2\|_{wh} < \infty$,

$$
\begin{aligned}
&\|\mathbf{E}\{\varphi_1(Y_k, M_k)\varphi_2(Y_{k+n}, M_{k+n})|Y_0 = y, M_0 = u\} \\
&\quad - \{\mathbf{E}\varphi_1(Y_k, M_k)|Y_0 = y, M_0 = u\} \\
&\qquad \times \{\mathbf{E}\varphi_2(Y_{k+n}, M_{k+n}|Y_0 = y, M_0 = u)\}\|_{wh} \leq r_1\gamma_1^n.
\end{aligned} \tag{4.14}
$$

Let $\tilde{\varphi}_1, \tilde{\varphi}_2$ be real-valued measurable functions on $(\mathcal{Y} \times \mathbf{M}) \times (\mathcal{Y} \times \mathbf{M})$. Denote $\varphi_1(z, v) = \mathbf{E}\{\tilde{\varphi}_1((z, v), (Y_1, M_1))|Y_0 = z, M_0 = v)\}$, and note that

$$
\begin{aligned}
&\mathbf{E}\{\tilde{\varphi}_1\big((Y_k, M_k), (Y_{k+1}, M_{k+1})\big)|Y_0 = y, M_0 = u\} \\
&\quad = \mathbf{E}\{\varphi_1(Y_k, M_k)|Y_0 = y, M_0 = u\}.
\end{aligned}
$$

The same proof as that of Theorem 16.1.5 of [46] can be used to show that there exist $r_1 > 0$ and $0 < \gamma_1 < 1$ such that, for all $y \in \mathcal{Y}, u \in \mathbf{M}, k \geq 0$ and $n \geq 1$ and for all measurable $\tilde{\varphi}_1, \tilde{\varphi}_2$ with $\|\sup_{z,v} \tilde{\varphi}_1^2((y, u), (z, v))\|_{wh} < \infty$ and $\|\sup_{z,v} \tilde{\varphi}_2^2((y, u), (z, v))\|_{wh} < \infty$,

$$
\begin{aligned}
&\big\|\mathbf{E}\{\tilde{\varphi}_1\big((Y_k, M_k), (Y_{k+1}, M_{k+1})\big) \\
&\qquad \times \tilde{\varphi}_2\big((Y_{k+n}, M_{k+n}), (Y_{k+n+1}, M_{k+n+1})\big)|Y_0 = y, M_0 = u\big\} \\
&\quad - \mathbf{E}\{\varphi(Y_k, M_k)|Y_0 = y, M_0 = u\big\}\mathbf{E}\{\varphi_2(Y_{k+n}, M_{k+n})|Y_0 = y, M_0 = u\}\big\|_{wh} \\
&\quad \leq r_1\gamma_1^{n-1}.
\end{aligned} \tag{4.15}
$$

To establish Edgeworth expansion for a Markovian iterated random functions system, we shall make use of (4.15) in conjunction with the following extension of Cramér (strongly nonlattice) condition:

$$
\inf_{|v| > \alpha} |1 - E_\pi\{\exp(iv' S_1(g)\}| > 0 \qquad \text{for all } \alpha > 0. \tag{4.16}
$$

In addition, we also assume the conditional Cramér (strongly nonlattice) condition ((2.5) on page 216 in [31]): There exists $\delta > 0$ such that, for all $m, n = 1, 2, \ldots$, $\delta^{-1} < m < n$, and all $\alpha \in \mathbf{R}^p$ with $|\alpha| \geq \delta$,

$$
\begin{aligned}
&E_\pi\big|E\big\{\exp\big(i\alpha'\big(g(M_{n-m}) + \cdots + g(M_{n+m})\big)\big) \\
&\qquad \big|(Y_{n-m}, M_{n-m}), \ldots, (Y_{n-1}, M_{n-1}), \\
&\qquad (Y_{n+1}, M_{n+1}), \ldots, (Y_{n+m}, M_{n+m}), (Y_{n+m+1}, M_{n+m+1})\big\}\big| \leq e^{-\delta}.
\end{aligned} \tag{4.17}
$$

Let

$$
\gamma = \int \mathbf{E}_{yu} g(M_1)\Pi(dy \times du)\big(= \lambda'(0)\big), \tag{4.18}
$$

and denote by $V = (\partial^2\lambda(\alpha)/\partial\alpha_i\,\partial\alpha_j|_{\alpha=0})_{1 \leq i,j \leq p}$ the Hessian matrix of λ at 0. By Lemma 2,

$$
\lim_{n\to\infty} n^{-1}\mathbf{E}_\nu\big\{\big(g(M_n) - n\gamma\big)\big(g(M_n) - n\gamma\big)'\big\} = V. \tag{4.19}
$$

Let $\psi_n(\alpha) = \mathbf{E}_\nu(e^{i\alpha' g(M_n)})$. Then by Lemma 2 and the fact that $\nu_\alpha \mathbf{Q}_\alpha h_1$ has continuous partial derivatives of order $r-2$ in some neighborhood of $\alpha = 0$, we have the Taylor series expansion of $\psi_n(\alpha/\sqrt{n})$ for $|\alpha/\sqrt{n}| \le \varepsilon$ (some sufficiently small positive number):

$$\psi_n(\alpha/\sqrt{n})\left\{1 + \sum_{j=1}^{r-2} n^{-j/2}\tilde{\pi}_j(i\alpha)\right\} e^{-\alpha' V\alpha/2} + o(n^{-(r-2)/2}), \tag{4.20}$$

where $\tilde{\pi}_j(i\alpha)$ is a polynomial in $i\alpha$ of degree $3j$ whose coefficients are smooth functions of the partial derivatives of $\lambda(\alpha)$ at $\alpha = 0$ up to the order $j+2$ and those of $\nu_\alpha \mathbf{Q}_\alpha h_1$ at $\alpha = 0$ up to the order j. Letting D denote the $p \times 1$ vector whose jth component is the partial differentiation operator D_j with respect to the jth coordinate, define the differential operator $\tilde{\pi}_j(-D)$. As in the case of sums of i.i.d. zero-mean random vectors (cf. [5]), we obtain an Edgeworth expansion for the "formal density" of the distribution of $g(M_n)$ by replacing the $\tilde{\pi}_j(i\alpha)$ and $e^{-\alpha' V\alpha/2}$ in (4.20) by $\tilde{\pi}_j(-D)$ and $\phi_V(y)$, respectively, where ϕ_V is the density function of the q-variate normal distribution with mean 0 and covariance matrix V. Throughout the sequel we let $\mathbf{P}_\nu$ denote the probability measure under which (Y_0, M_0) has initial distribution ν.

THEOREM 2. *Let $\{(Y_n, M_n), n \ge 0\}$ be the MIRFS defined in* (2.1) *satisfying Assumption* K, *such that the induced Markov chain $\{(Y_n, M_n), n \ge 0\}$, with transition probability kernel* (3.2), *is irreducible, aperiodic and Harris recurrent. Assuming $g \in \mathcal{L}^r(Q)$ for some $r > 2$,* (4.16) *and* (4.17) *hold. Let $\phi_{j,V} = \tilde{\pi}_j(-D)\phi_V$ for $j = 1, \dots, r-2$. For $0 < a \le 1$ and $c > 0$, let $\mathcal{B}_{a,c}$ be the class of all Borel subsets B of $\mathbf{R}^p$ such that $\int_{(\partial B)^\varepsilon} \phi_V(y)\,dy \le c\varepsilon^a$ for every $\varepsilon > 0$, where ∂B denotes the boundary of B and $(\partial B)^\varepsilon$ denotes its ε-neighborhood. Then*

$$\begin{aligned}\sup_{B \in \mathcal{B}_{a,c}} & \left| \mathbf{P}_\nu\{(S_n - n\gamma)/\sqrt{n} \in B\} - \int_B \left\{\phi_V(y) + \sum_{j=1}^{r-2} n^{-j/2}\phi_{j,V}(y)\right\} dy \right| \\ & = o(n^{-(r-2)/2}).\end{aligned} \tag{4.21}$$

A proof of Theorem 2 is given in the Appendix.

Note that under weaker moment conditions, and an alternative condition of (4.16) and (4.17) (see Condition 1 of [42]) Lahiri [42] proved the asymptotic expansions for sums of weakly dependent random vectors.

Letting $r = 2$ in Theorem 2, we have the following:

COROLLARY 1. *With the same notation and assumptions as in Theorem* 2, *then*

$$\frac{1}{\sqrt{n}}(S_n - n\gamma) \longrightarrow N(0, \Sigma) \qquad \textit{in distribution},$$

where the variance–covariance matrix

$$\Sigma = \left(\frac{\partial^2 \lambda(\alpha)}{\partial \alpha_i \, \partial \alpha_j} \bigg|_{\alpha=0} \right)_{i,j=1,\dots,p}. \tag{4.22}$$

In statistical applications one often works with $\mathbf{g}(n^{-1}S_n)$ instead of $S_n = \sum_{k=1}^n g(M_k)$, where $\mathbf{g}: \mathbf{R}^p \to \mathbf{R}^q$ is sufficiently smooth in some neighborhood of the mean $\gamma := (\gamma_1, \dots, \gamma_p)$. Denote $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_q)$ with each $\mathbf{g}_i$, $1 \le i \le q$, a real-valued function on $\mathbf{R}^p$. For the case of a sum of i.i.d. random variables, Bhattacharya and Ghosh [4] made use of the Edgeworth expansion of the distribution of $(S_n - n\gamma)/\sqrt{n}$ to derive an Edgeworth expansion of the distribution of $\sqrt{n}\{\mathbf{g}(n^{-1}S_n) - \mathbf{g}(\gamma)\}$. Making use of Theorem 2 and a straightforward extension of their argument, we can generalize their result to the case where S_n is the partial sum of a Markovian iterated random functions system.

THEOREM 3. *Under the same assumptions as in Theorem 2, suppose that $\mathbf{g}: \mathbf{R}^p \to \mathbf{R}^q$ has continuous partial derivatives of order r in some neighborhood of γ. Let $J_{\mathbf{g}} = (D_j \mathbf{g}_i(\gamma))_{1 \le i \le q, 1 \le j \le p}$ be the $q \times p$ Jacobian matrix and let $V(\mathbf{g}) = J_{\mathbf{g}} V J'_{\mathbf{g}}$. Then*

$$\begin{aligned} \sup_{B \in \mathscr{B}_{a,c}} \bigg| & \mathbf{P}_\nu\{\sqrt{n}(\mathbf{g}(n^{-1}S_n) - \mathbf{g}(\gamma)) \in B\} \\ & - \int_B \left\{ \phi_{V(\mathbf{g})}(y) + \sum_{j=1}^{r-2} n^{-j/2} \phi_{j,V,\mathbf{g}}(y) \right\} dy \bigg| \\ & = o(n^{-(r-2)/2}), \end{aligned} \tag{4.23}$$

where $\phi_{j,V,\mathbf{g}} = \tilde{\pi}_{j,\mathbf{g}}(-D)\phi_V$ and $\tilde{\pi}_{j,\mathbf{g}}(y)$ is a polynomial in $y (\in \mathbf{R}^p)$ whose coefficients are smooth functions of the partial derivatives of $\lambda(\alpha)$ at $\alpha = 0$ up to order $j+2$ and those of $\nu_\alpha \mathbf{Q}_\alpha h_1$ at $\alpha = 0$ up to order j together with those of $\mathbf{g}$ at μ up to order $j+1$.

In the next theorem we consider $p = 1$.

THEOREM 4. *Under the same assumptions as in Theorem 2, assume $g \in \mathcal{L}^r(Q)$ for some $r > 2$. Then*

$$\frac{1 - \mathbf{P}_\nu\{(S_n - n\gamma)/\sqrt{n} \le t\}}{1 - \Phi(t)} = \exp(t^3/\sqrt{n})\varphi(t/\sqrt{n})\left(1 + O\left(\frac{t}{\sqrt{n}}\right)\right) \tag{4.24}$$

and

$$\frac{\mathbf{P}_\nu\{(S_n - n\gamma)/\sqrt{n} \le -t\}}{\Phi(-t)} = \exp(-t^3/\sqrt{n})\varphi(-t/\sqrt{n})\left(1 + O\left(\frac{t}{\sqrt{n}}\right)\right), \tag{4.25}$$

where $\Phi(t)$ is the standard normal distribution, and $\varphi(t)$ is a power series which converges for t sufficiently small in absolute value.

Theorem 4 states the moderate deviations results for the distribution of an MIRFS, which will be used to prove Edgeworth expansion for the MLE in Section 5. Since the proof is a straightforward generalization of Theorem 6 in [47], it will not be repeated here.

5. Efficient likelihood estimation. For a given state space model defined in (2.1) which involves several parameters $\theta = (\theta_1, \dots, \theta_q)$, the estimation problem we consider in this section is the case of estimating one of the parameters at a time; the other parameters play the role of nuisance parameters. The true parameter is denoted by θ_0. Recall $p_n = p_n(\xi_0, \xi_1, \dots, \xi_n; \theta)$ defined as (2.3). When $\partial \log p_n / \partial \theta$ exists, one can seek solutions of the likelihood equations

$$\frac{\partial \log p_n}{\partial \theta} = 0. \tag{5.1}$$

In the following, we denote E_x^θ as the expectation defined under $P^\theta(\cdot, \cdot)$ in (2.1) with initial state $X_0 = x$, and $E_{(x,s)}^\theta$ as the expectation defined under $P^\theta(\cdot, \cdot)$ in (2.1) with initial state $X_0 = x, \xi_0 = s$. The following conditions will be used throughout the rest of this paper.

C1. For given $\theta \in \Theta$, the Markov chain $\{(X_n, \xi_n), n \geq 0\}$ defined in (2.1) and (2.2) is aperiodic, irreducible, and satisfies (4.1) and (4.2) with weight function $w(\cdot)$. Assume $0 < p_\theta(x, y) < \infty$ for all $x, y \in \mathcal{X}$, and $0 < \sup_{x \in \mathcal{X}} f(s_1; \theta | x, s_0) < \infty$, for all $s_0, s_1 \in \mathbf{R}^d$. Denote $g_\theta(s_0, \xi_1) = \sup_{x_0 \in \mathcal{X}} \int p_\theta(x_0, x_1) f(\xi_1; \theta | x_1, s_0) \times m(dx_1)$. Furthermore, we assume that there exists $p \geq 1$ as in Assumption K2 such that

$$\sup_{(x_0, s_0) \in \mathcal{X} \times \mathbf{R}^d} E_{(x_0, s_0)}^\theta \left\{ \log\left(g_\theta(s_0, \xi_1)^p \frac{w(X_p, \xi_p)}{w(x_0, s_0)} \right) \right\} < 0, \tag{5.2}$$

$$\sup_{(x_0, s_0) \in \mathcal{X} \times \mathbf{R}^d} E_{(x_0, s_0)}^\theta \left\{ g_\theta(s_0, \xi_1) \frac{w(X_1, \xi_1)}{w(x_0, s_0)} \right\} < \infty. \tag{5.3}$$

C2. The true parameter θ_0 is an interior point of Θ. For all $x \in \mathcal{X}$, $s_0, s_1 \in \mathbf{R}^d$, $\theta \in \Theta \subset \mathbf{R}^q$, and for $i, j, k = 1, \dots, q$, the partial derivatives

$$\frac{\partial f(s_0; \theta | x)}{\partial \theta_i}, \qquad \frac{\partial^2 f(s_0; \theta | x)}{\partial \theta_i \, \partial \theta_j}, \qquad \frac{\partial^3 f(s_0; \theta | x)}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k} \qquad \text{exist,}$$

as well as the partial derivatives

$$\frac{\partial f(s_1; \theta | x, s_0)}{\partial \theta_i}, \qquad \frac{\partial^2 f(s_1; \theta | x, s_0)}{\partial \theta_i \, \partial \theta_j}, \qquad \frac{\partial^3 f(s_1; \theta | x, s_0)}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k},$$

and for all $x, y \in \mathcal{X}$, $\theta \to p_\theta(x, y)$ and $\theta \to \pi_\theta(x)$ have twice continuous derivatives in some neighborhood $N_\delta(\theta_0) := \{\theta : |\theta - \theta_0| < \delta\}$ of θ_0.

C3.

$$\int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} \left|\frac{\partial \pi_\theta(x)}{\partial \theta_i}\right| m(dx) < \infty, \qquad \int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} \left|\frac{\partial^2 \pi_\theta(x)}{\partial \theta_i\, \partial \theta_j}\right| m(dx) < \infty,$$

and for all $x \in \mathcal{X}$, $i, j = 1, \dots, q$,

$$\int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} \left|\frac{\partial p_\theta(x, y)}{\partial \theta_i}\right| m(dy) < \infty, \qquad \int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} \left|\frac{\partial^2 p_\theta(x, y)}{\partial \theta_i\, \partial \theta_j}\right| m(dy) < \infty.$$

C4. For all $x \in \mathcal{X}$, $s_0 \in \mathbf{R}^d$ and $\theta \in \Theta$,

$$E_x^\theta \left|\frac{\partial f(\xi_0; \theta|x)}{\partial \theta_i}\right| < \infty, \qquad E_x^\theta \left|\frac{\partial^2 f(\xi_0; \theta|x)}{\partial \theta_i\, \partial \theta_j}\right| < \infty,$$

$$E_{(x,s_0)}^\theta \left|\frac{\partial f(\xi_1; \theta|x, s_0)}{\partial \theta_i}\right| < \infty, \qquad E_{(x.s_0)}^\theta \left|\frac{\partial^2 f(\xi_1; \theta|x, s_0)}{\partial \theta_i\, \partial \theta_j}\right| < \infty.$$

Furthermore, we assume that, for all $x \in \mathcal{X}$, $s_0 \in \mathbf{R}^d$ and uniformly for $\theta \in N_\delta(\theta_0)$,

$$\left|\frac{\partial^3 \log f(\xi_0; \theta|x)}{\partial \theta_i\, \partial \theta_j\, \partial \theta_k}\right| < H_{ijk}(x, \xi_0), \qquad \left|\frac{\partial^3 \log f(\xi_1; \theta|x, s_0)}{\partial \theta_i\, \partial \theta_j\, \partial \theta_k}\right| < G_{ijk}\big((x, s_0), \xi_1\big),$$

where H_{ijk} and G_{ijk} are such that $E_x^{\theta_0} H_{ijk}(x, \xi_0) < \infty$ and $E_{(x,s_0)}^{\theta_0} G_{ijk}((x, s_0), \xi_1) < \infty$, for all $i, j, k = 1, \dots, q$ and for all $x \in \mathcal{X}$, $s_0 \in \mathbf{R}^d$.

C5.

$$\sup_{x \in \mathcal{X}} E_x^{\theta_0} \left(\sup_{|\theta - \theta_0| < \delta} \sup_{y, z \in \mathcal{X}} \frac{f(\xi_0; \theta|y) f(\xi_1; \theta|y, \xi_0)}{f(\xi_0; \theta|z) f(\xi_1; \theta|z, \xi_0)} \right)^2 < \infty.$$

C6. The equality

$$p_n(\xi_0, \xi_1, \dots, \xi_n; \theta) = p_n(\xi_0, \xi_1, \dots, \xi_n; \theta')$$

holds P-almost surely, for all nonnegative n, if and only if $\theta = \theta'$.

C7. For all $x, y \in \mathcal{X}$, $\theta \to p_\theta(x, y)$, $\theta \to \pi_\theta(x)$ and $\theta \to \varphi_x(\theta)$, are continuous, and $\theta \to f(s_0; \theta|x)$, as well as $\theta \to f(s_1; \theta|x, s_0)$, are continuous for all $x \in \mathcal{X}$ and $s_0, s_1 \in \mathbf{R}^d$. Furthermore, for all $x \in \mathcal{X}$ and $s_0, s_1 \in \mathbf{R}^d$, $f(s_0; \theta|x) \to 0$ and $f(s_1; \theta|x, s_0) \to 0$, as $|\theta| \to \infty$.

C8. $E_x^{\theta_0} |\log(f(\xi_0; \theta_0|x) f(\xi_1; \theta_0|x, \xi_0))| < \infty$ for all $x \in \mathcal{X}$.

C9. For each $\theta \in \Theta$, there is $\delta > 0$ such that, for all $x \in \mathcal{X}$,

$$E_x^{\theta_0} \Big(\sup_{|\theta' - \theta| < \delta} \big[\log\big(f(\xi_0; \theta'|x) f(\xi_1; \theta'|x, \xi_0)\big)\big]^+ \Big) < \infty,$$

where $a^+ = \max\{a, 0\}$. And there is a $b > 0$ such that, for all $x \in \mathcal{X}$,

$$E_x^{\theta_0} \Big(\sup_{|\theta'| > b} \big[\log\big(f(\xi_0; \theta'|x) f(\xi_1; \theta'|x, \xi_0)\big)\big]^+ \Big) < \infty.$$

REMARK 2. (a) Condition C1 is the w-uniform ergodicity condition for the underlying Markov chain, which is considerably weaker than the uniformly recurrent condition A1 of [39], and that of [14]. Furthermore, we impose conditions (5.2) and (5.3) to guarantee that the induced Markovian iterated random functions system satisfies Assumptions K2 and K3 in Section 4.

(b) To have better understanding of these properties, we first consider a simple state space model $X_n = \alpha X_{n-1} + \varepsilon_n, \xi_n = X_n + \eta_n$, where $|\alpha| < 1$, ε_n and η_n are i.i.d. standard normal random variables, and they are mutually independent. Since ξ_n are independent for given X_n, the weight function w depends on X_0 only and we have $w(x) = |x| + 1$. Note that $\mathcal{X} = \mathbf{R}$. Denote $b = (1 - |\alpha|^p)/(1 - |\alpha|)$. Observe that

$$\begin{aligned}
&\sup_{x \in \mathbf{R}} \int_{-\infty}^{\infty} \frac{\exp\{-(y - \alpha x)^2/2\}}{\sqrt{2\pi}} \frac{\exp\{-(s - y)^2/2\}}{\sqrt{2\pi}}\, dy \\
&\quad = \sup_{x \in \mathbf{R}} \frac{\sqrt{1/2}}{\sqrt{2\pi}} \exp\{-(\alpha x - s)^2/4\} \\
&\qquad \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1/2)}} \exp\{-(y - (\alpha x + s)/2)^2/2(1/2)\}\, dy \\
&\quad = \frac{\sqrt{1/2}}{\sqrt{2\pi}} \sup_{x \in \mathbf{R}} \exp\{-(\alpha x - s)^2/4\} = \frac{1}{\sqrt{4\pi}}.
\end{aligned}$$

A simple calculation leads to

$$\begin{aligned}
&\sup_{(x_0, s_0) \in \mathbf{R} \times \mathbf{R}} E^{\alpha}_{(x_0, s_0)} \left\{ \log\left(g(s_0, \xi_1)^p \frac{w(X_p, \xi_p)}{w(x_0, s_0)} \right) \right\} \\
&\quad < \log \sup_{x_0 \in \mathbf{R}} E^{\alpha}_{x_0} \left\{ \frac{|\alpha^p x_0 + \sum_{k=0}^{p-1} \alpha^k \varepsilon_{p-k}| + 1}{(4\pi)^{p/2}(|x_0| + 1)} \right\} \\
&\quad \le \log \sup_{x_0 \in \mathbf{R}} \left\{ \frac{|\alpha^p x_0| + E^{\alpha}_{x_0} |\sum_{k=0}^{p-1} \alpha^k \varepsilon_{p-k}| + 1}{(4\pi)^{p/2}(|x_0| + 1)} \right\} \\
&\quad = \log \sup_{x_0 \in \mathbf{R}} \left\{ \frac{|\alpha^p x_0| + 2b/\sqrt{2\pi} + 1}{(4\pi)^{p/2}(|x_0| + 1)} \right\} < 0.
\end{aligned} \tag{5.4}$$

This implies that (5.2) holds. By using the same argument, we see (5.3) holds.

Next, we consider the case that ε_n and η_n are i.i.d. double exponential(1) random variables. Observe that

$$\begin{aligned}
&\sup_{x \in \mathbf{R}} \int_{-\infty}^{\infty} \frac{\exp\{-|y - \alpha x|\}}{\sqrt{2}} \frac{\exp\{-|s - y|\}}{\sqrt{2}}\, dy \\
&\quad = \frac{1}{4} \sup_{x \in \mathbf{R}} \big((1 + |\alpha x - s|) \exp\{-|\alpha x - s|\}\big) = \frac{1}{4}.
\end{aligned}$$

By making use of the same argument as in (5.4), we see that (5.2) and (5.3) hold. The extension to $\xi_n = \beta_{X_n}\xi_{n-1} + \eta_n$, studied in Remark 1(b), is straightforward and will not be repeated here. Other practical used models of the Markov-switching model, ARMA models, (G)ARCH models and SV models will be given in Section 6.

(c) Note that the mean contraction property $\mathbf{E}\log L_1 < 0$ is not satisfied in the above examples. Instead of applying Theorem 1 directly, we will explore the special structure of the likelihood function in Lemma 4 below, such that $\{((X_n, \xi_n), M_n), n \geq 0\}$ is an irreducible, aperiodic and Harris recurrent Markov chain. Hence, we can apply Theorem 1 for the Markovian iterated functions system on $\mathbf{M}$ induced from (2.4)–(2.7).

(d) C2–C4 are standard smoothness conditions. C5 is the technical condition for the existence of the Fisher information to be defined in (5.9) below. C8 and C9 are integrability conditions that will be used to prove strong consistency of the MLE. Condition C6 is the identifiability condition for state space models. That is, the family of mixtures of $\{f(\xi_1;\theta|x,\xi_0) : \theta \in \Theta\}$ is identifiable. This condition will be used to prove strong consistency of the MLE. Although it is difficult to check this condition in a general state space model, in many models of interest the parameter itself is identifiable only up to a permutation of states such as a finite state hidden Markov model with normal distributions. A sufficient condition for the identifiable issue can be found in Theorem 1 of [14]. See also the paper by Itô, Amari and Kobayashi [38] for necessary and sufficient conditions in the case that the state space is finite and ξ_i is a deterministic function of X_i.

(e) When the state space of the Markov chain $\{X_n, n \geq 0\}$ is finite, and the observations ξ_n are conditionally independent, this reduces to the so-called hidden Markov model. It is easy to see that condition C1 implies (A1) by choosing $w(x) = 1$, and conditions C2–C4 reduce to (A2), (A3) and (A5) of [7]. Conditions C6–C9 reduce to conditions C1–C6 in [44]. We will discuss condition C5 in Remark 3 after Lemma 5.

Let $\{(X_n, \xi_n), n \geq 0\}$ be the Markov chain defined in (2.1) and (2.2). Recall from (2.8) that the log likelihood can be written as

$$\begin{aligned} l(\theta) &= \log p_n(\xi_1, \ldots, \xi_n; \theta) = \log \|\mathbf{P}_\theta(\xi_n) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\| \\ &= \log \frac{\|\mathbf{P}_\theta(\xi_n) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|}{\|\mathbf{P}_\theta(\xi_{n-1}) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|} \\ &\quad + \cdots + \log \frac{\|\mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|}{\|\mathbf{P}_\theta(\xi_0)\pi\|}. \end{aligned} \tag{5.5}$$

For each n, denote

$$M_n := \mathbf{P}_\theta(\xi_n) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0) \tag{5.6}$$

as the Markovian iterated random functions system on $\mathbf{M}$ induced from (2.4)–(2.7). Then $\{((X_n,\xi_n),M_n), n\ge 0\}$ is a Markov chain on the state space $(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}$, with transition probability kernel $\mathbf{P}_\theta$ defined as in (3.2). Let Π_θ be the stationary distribution of $\{((X_n,\xi_n),M_n), n\ge 0\}$ defined in Theorem 1(iii). Then the log-likelihood function $l(\theta)$ can be written as $S_n := \sum_{k=1}^n g(M_{k-1},M_k)$ with

$$g(M_{k-1},M_k) := \log \frac{\|\mathbf{P}_\theta(\xi_k)\circ\cdots\circ\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)\pi\|}{\|\mathbf{P}_\theta(\xi_{k-1})\circ\cdots\circ\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)\pi\|}. \tag{5.7}$$

In order to apply Theorems 1–4, we need to check that the Markovian iterated random functions system satisfies Assumption K, and the induced Markov chain is aperiodic, irreducible and Harris recurrent. For this purpose, we need to define a suitable metric on the space $\mathbf{M}$, which has been defined in (2.4). First, we add a further condition on $\mathbf{M}$ to have

$$\mathbf{M} = \left\{h | h : \mathcal{X}\to\mathbf{R}^+ \text{ is } m\text{-measurable}, \int h(x)m(dx) < \infty \text{ and } \sup_{x\in\mathcal{X}} h(x) < \infty\right\}.$$

For convenience of notation, we still use the notation $\mathbf{M}$, and will use h to represent an element in $\mathbf{M}$, which is different from the notation u used in Sections 3 and 4. We define the variation distance between any two elements h_1, h_2 in $\mathbf{M}$ by

$$d(h_1,h_2) = \sup_{x\in\mathcal{X}} |h_1(x) - h_2(x)|. \tag{5.8}$$

Note that $(\mathbf{M}, d)$ is a complete metric space with Borel σ-algebra $\mathcal{B}(\mathbf{M})$, but it is not separable. Thus, Theorems 1–4 do not apply. However, rather than deal with the measure-theoretic technicalities created by an inseparable space, we can apply the results developed in Section 7 of [13] for a direct argument of convergence. Therefore, Theorems 1–4 still hold under the regularity conditions.

In order to describe our main results, we need the following lemmas first. Their proofs are given in Section 7.

LEMMA 3. *Assume* C1–C5 *hold or* C1, C6–C9 *hold. Then for each* $\theta\in\Theta$ *and* $j = 1,\dots,n$, *the random functions* $\mathbf{P}_\theta(\xi_0)$ *and* $\mathbf{P}_\theta(\xi_j)$, *defined in* (2.5) *and* (2.6), *from* $(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}$ *to* $\mathbf{M}$ *are Lipschitz continuous in the second argument, and the Markovian iterated random functions system* (2.4)–(2.7) *satisfies Assumption* K. *Furthermore, the function* g *defined in* (5.7) *belongs to* $\mathcal{L}^r(Q)$ *for any* $r > 0$.

For each $\theta\in\Theta$, recall that $\{((X_n,\xi_n),M_n), n\ge 0\}$ is a Markov chain induced by the Markovian iterated random functions system (2.4)–(2.7) on the state space $(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}$.

LEMMA 4. *Assume* C1–C5 *hold or* C1, C6–C9 *hold. Then for each* $\theta\in\Theta$, $\{((X_n,\xi_n),M_n), n\ge 0\}$ *is an aperiodic*, $(m\times Q\times Q)$*-irreducible and Harris recurrent Markov chain.*

LEMMA 5. *Assume* C1–C5 *hold. Then the Fisher information matrix*

$$\mathbf{I}(\theta) = (I_{ij}(\theta)) \tag{5.9}$$
$$= \left(\mathbf{E}_{\Pi}^{\theta}\left[\left(\frac{\partial \log \|\mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|}{\partial \theta_i}\right) \times \left(\frac{\partial \log \|\mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|}{\partial \theta_j}\right)\right]\right)$$

is positive definite for θ *in a neighborhood* $N_\delta(\theta_0)$ *of* θ_0. *Recall that* $\mathbf{E}_{\Pi}^{\theta} := \mathbf{E}_{\Pi}$ *is defined as the expectation under* $\mathbf{P}_{\Pi}$ *in* (3.2).

REMARK 3. Note that the Fisher information (5.9) is defined as the expected value under the stationary distribution Π^θ of the Markov chain $\{((X_n, \xi_n), M_n), n \geq 0\}$. It is worth mentioning that only ξ_n appears in M_n, in which it reflects the nature of state space models.

When the state space $\mathcal{X}$ is finite, and the random variables ξ_n are conditionally independent for given X_n, let $H := H(\xi_1, \xi_0, \xi_{-1}, \ldots) = \sum_{m=-\infty}^{1} H_m(\xi_1, \xi_0, \ldots)$, where

$$\begin{aligned} H_m(\xi_1, \xi_0, \ldots) := {} & E^{\theta^0}\left\{\frac{\partial \log f(\xi_m; \theta | X_m)}{\partial \theta}\Big|\xi_1, \xi_0, \ldots\right\} \\ & - E^{\theta^0}\left\{\frac{\partial \log f(\xi_m; \theta | X_m)}{\partial \theta}\Big|\xi_0, \xi_{-1}, \ldots\right\} \\ & + E^{\theta^0}\left\{\frac{\partial \log p_\theta(X_m, X_{m+1})}{\partial \theta}\Big|\xi_1, \xi_0, \ldots\right\} \\ & - E^{\theta^0}\left\{\frac{\partial \log p_\theta(X_m X_{m+1})}{\partial \theta}\Big|\xi_0, \xi_{-1}, \ldots\right\}. \end{aligned}$$

Under their Assumptions 1–4, Bickel and Ritov [6] showed that $H \in \mathcal{L}^2(P^{\theta^0})$ and defined $\mathbf{I}_H(\theta^0) := E^{\theta^0}\{HH^t\}$. They also showed that

$$\lim_{n\to\infty} \frac{1}{n} E^{\theta^0}\left(\left(\frac{\partial \log \|T_n\pi\| |_{\theta=\theta^0}}{\partial \theta}\right)\left(\frac{\partial \log \|T_n\pi\| |_{\theta=\theta^0}}{\partial \theta}\right)^t\right) = \mathbf{I}_H(\theta^0).$$

In this paper we represent the log likelihood function of an additive functional of the Markov chain $\{((X_n, \xi_n), M_n), n \geq 0\}$ in (5.7), and then apply the strong law of large numbers for Markovian iterated random functions given in Theorem 1(iv) to have, with probability 1,

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \log \|\mathbf{P}_\theta(\xi_n) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\| = -I_{ij}(\theta).$$

Hence, under Assumptions 1–4 of [6], $\mathbf{I}(\theta)$ is well defined and is equal to $\mathbf{I}_H(\theta)$. The moment condition in Assumption 4 of [6] can be relaxed to the following: there exists a $\delta > 0$ with $\rho_0(\xi) := \sup_{|\theta-\theta^0|<\delta} \max_{x,y\in\mathcal{X}} \frac{f(\xi;\theta|x)}{f(\xi;\theta|y)}$, such that $\sup_{x\in\mathcal{X}} P^{\theta^0}\{\rho_0(\xi_1) = \infty | X_0 = x\} < 1$; see [7].

LEMMA 6. *Assume* C1–C5 *hold. Let* $l_j'(\theta_0) = \partial l(\theta)/\partial\theta_j|_{\theta=\theta_0}$. *Then, as* $n \to \infty$,

$$\frac{1}{\sqrt{n}}(l_j'(\theta_0))_{j=1,\dots,q} \longrightarrow N\big(0, \mathbf{I}(\theta_0)\big) \qquad \text{in distribution.} \tag{5.10}$$

THEOREM 5. *Assume* C1–C5 *hold. Then there exists a sequence of solutions* $\hat{\theta}_n$ *of* (5.1) *such that* $\hat{\theta}_n \to \theta_0$ *in probability. Furthermore,* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is asymptotically normally distributed with mean zero and variance–covariance matrix* $\mathbf{I}^{-1}(\theta_0)$.

Since the proof of Theorem 5 follows a standard argument, we will not give it here.

COROLLARY 2. *Under the assumptions of Theorem* 5, *if the likelihood equation has a unique root for each* n *and all* $\xi_1, \dots, \xi_n$, *then there is a consistent sequence of estimators* $\hat{\theta}_n$ *of the unknown parameters* θ_0.

Next, we prove strong consistency of the MLE when the log likelihood function is integrable. A crucial step is to give an appropriate definition of the Kullback–Leibler information for state space models, so that we can apply Theorem 1 to have a standard argument of strong consistency for the MLE. Here, we define the Kullback–Leibler information as

$$\begin{aligned} K(\theta_0, \theta) &= \mathbf{E}_{\Pi}^{\theta_0}\left(\log \frac{\|\mathbf{P}_{\theta_0}(\xi_1) \circ \mathbf{P}_{\theta_0}(\xi_0)\pi_{\theta_0}\|}{\|\mathbf{P}_{\theta}(\xi_1) \circ \mathbf{P}_{\theta}(\xi_0)\pi_{\theta}\|}\right) \\ &:= \int \log \frac{\|\mathbf{P}_{\theta_0}(\xi_1) \circ \mathbf{P}_{\theta_0}(\xi_0)\pi_{\theta_0}\|}{\|\mathbf{P}_{\theta}(\xi_1) \circ \mathbf{P}_{\theta}(\xi_0)\pi_{\theta}\|} \Pi\big(d(x,\xi) \times d\pi_{\theta_0}\big). \end{aligned} \tag{5.11}$$

THEOREM 6. *Assume that* C1, C6–C9 *hold and let* $\hat{\theta}_n$ *be the MLE based on* n *observations* $\xi_0, \xi_1, \dots, \xi_n$. *Then* $\hat{\theta}_n \longrightarrow \theta_0$ P^{θ_0}*-a.s. as* $n \to \infty$.

Since the proof of Theorem 6 follows a standard argument, we will not give it here.

To derive the Edgeworth expansion for the MLE, we need to define the following notation and assumptions first. For nonnegative integral vectors $\nu = (\nu^{(1)}, \dots, \nu^{(q)})$, write $|\nu| = \nu^{(1)} + \cdots + \nu^{(q)}$, $\nu! = \nu^{(1)}! \cdots \nu^{(q)}!$, and let $D^\nu = (D_1)^{\nu^{(1)}} \cdots (D_q)^{\nu^{(q)}}$ denote the νth derivative with respect to θ. Suppose assumptions C2, C3, C4 and C5 are strengthened so that there exists $r \geq 3$, as follows.

C2′. The true parameter θ_0 is an interior point of Θ. For all $x \in \mathcal{X}$, $s_0, s_1 \in \mathbf{R}^d$, $\theta \in \Theta \subset \mathbf{R}^q$, the partial derivatives

$$D^1 f(s_0; \theta|x), \qquad D^2 f(s_0; \theta|x), \dots, D^r f(s_0; \theta|x),$$

as well as the partial derivatives

$$D^1 f(s_1; \theta|x, s_0), \qquad D^2 f(s_1; \theta|x, s_0), \dots, D^r f(s_1; \theta|x, s_0),$$

and for all $x, y \in \mathcal{X}$, $\theta \to p_\theta(x, y)$ and $\theta \to \pi_\theta(x)$ have $r - 1$ continuous derivatives in some neighborhood $N_\delta(\theta_0) := \{\theta : |\theta - \theta_0| < \delta\}$ of θ_0.

C3′.

$$\int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} |D^1 \pi_\theta(x)| m(dx) < \infty, \dots, \int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} |D^{r-1} \pi_\theta(x)| m(dx) < \infty,$$

and for all $x \in \mathcal{X}$,

$$\int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} |D^1 p_\theta(x, y)| m(dy) < \infty, \dots, \int_{\mathcal{X}} \sup_{\theta \in N_\delta(\theta_0)} |D^{r-1} p_\theta(x, y)| m(dy) < \infty.$$

C4′. For all $x \in \mathcal{X}$, $s_0 \in \mathbf{R}^d$ and $\theta \in \Theta$,

$$E_x^\theta |D^\nu f(\xi_0; \theta|x)|^r < \infty, \qquad E_{(x,s_0)}^\theta |D^\nu f(\xi_1; \theta|x, s_0)|^r < \infty,$$

for $1 \le |\nu| \le r$, and

$$E_x^\theta \Bigl(\sup_{\theta \in N_\delta(\theta_0)} |D^\nu f(\xi_0; \theta|x)|^r \Bigr) < \infty,$$

$$E_{(x,s_0)}^\theta \Bigl(\sup_{\theta \in N_\delta(\theta_0)} |D^\nu f(\xi_1; \theta|x, s_0)|^r \Bigr) < \infty,$$

for $|\nu| = r + 1$.

C5′.

$$\sup_{x \in \mathcal{X}} E_x^{\theta_0} \biggl(\sup_{|\theta - \theta_0| < \delta} \sup_{y, z \in \mathcal{X}} \frac{f(\xi_0; \theta|y) f(\xi_1; \theta|y, \xi_0)}{f(\xi_0; \theta|z) f(\xi_1; \theta|z, \xi_0)} \biggr)^r < \infty.$$

We will assume conditions (4.16) and (4.17) hold for $Z_j^{(\nu)} := D^\nu \log p_1(\xi_0, \xi_1; \theta_0)$, $1 \le |\nu| \le r$. Let $Z_j := \{Z_j^{(\nu)} : 1 \le |\nu| \le r\}$ be p-dimensional random vectors for $j \ge 1$, where p is the number of all distinct multi-indices ν, $1 \le |\nu| \le r$. In the following, denote $\bar{Z} = (1/n) \sum_{k=1}^n Z_k$.

Use a standard argument involving the sign change of a continuous function, or a fixed point theorem in the multi-parameter case (cf. [4]), to prove that the likelihood equation has a solution which converges in probability to θ_0. Note that the following notation is interpreted in the multi-dimensional sense. Applying the moderate deviation result on $\bar{Z}$ in Theorem 4, it is possible to ensure that, with P^{θ_0}-probability $1 - o(n^{-1})$, $\hat{\theta}_n$ satisfies the likelihood equation and lies on $(\theta_0 \pm$

$\log n/\sqrt{n}$). It is this solution we take as our $\hat{\theta}_n$. If the likelihood equation has multiple roots, assume we have a consistent estimator T_n such that T_n lies in ($\theta_0 \pm \log n/\sqrt{n}$) with P^{θ_0}-probability $1 - o(n^{-1})$. In this case, we may take the solution nearest to T_n. By the preceding reasoning, this solution, which is identifiable from the sample, will lie in ($\theta_0 \pm \log n/\sqrt{n}$) with P^{θ_0}-probability $1 - o(n^{-1})$.

Clearly, with $\hat{\theta}_n$ as above, with probability $1 - o(n^{-1})$,

$$0 = \bar{Z}^{(e_s)} + \sum_{|\nu|=1}^{r-1} \frac{1}{\nu!} \bar{Z}^{(e_s+\nu)} (\hat{\theta}_n - \theta_0)^\nu + R_{n,s}(\hat{\theta}_n), \qquad 1 \le s \le q, \tag{5.12}$$

where e_s has 1 as the sth coordinate and zeros otherwise.

We rewrite equation (5.12) as

$$0 = A(\bar{Z}, \hat{\theta}_n) + R_n. \tag{5.13}$$

Note $0 = A(\gamma(\theta_0), \theta_0)$ and $\frac{\partial A}{\partial \theta}|_{\gamma(\theta_0),\theta_0} = -(\text{Fisher information}) \neq 0$.

Hence, by the implicit function theorem, there are a neighborhood N of γ and q uniquely defined real-valued infinitely differentiable functions $\mathbf{g}_i$ $(1 \le i \le q)$ on N such that $\theta = \mathbf{g}(z) = (\mathbf{g}_1(z), \dots, \mathbf{g}_q(z))$ satisfies (5.13). This implies, with probability $1 - o(n^{-1})$, $|\hat{\theta}_n - \theta_0| \le K(\log n/\sqrt{n})^4$.

To derive the asymptotic expansion of $P^{\theta_0}\{\sqrt{n}(\hat{\theta}_n - \theta_0) \in B\}$, note that $\hat{\theta}_n = \mathbf{g}(n^{-1} Z_n)$, where $\mathbf{g}: \mathbf{R}^p \to \mathbf{R}^q$ is sufficiently smooth in some neighborhood of γ. For the case of i.i.d. ξ_n, Bhattacharya and Ghosh [4] made use of the Edgeworth expansion of the distribution of $(S_n - n\gamma)/\sqrt{n}$ to derive an Edgeworth expansion of the distribution of $\sqrt{n}\{\mathbf{g}(n^{-1} S_n) - \mathbf{g}(\gamma)\}$. Making use of Theorem 4 and a straightforward extension of their argument, we can generalize their result to have the following theorem.

THEOREM 7. *Assume* C1, C2′–C5′ *hold for some* $r \ge 3$. *Assume* (4.16) *and* (4.17) *hold. Let* $J_{\mathbf{g}} = (D_j \mathbf{g}_i(\gamma))_{1 \le i \le q, 1 \le j \le p}$ *be the* $q \times p$ *Jacobian matrix and let* $V(\mathbf{g}) = J_{\mathbf{g}} V J_{\mathbf{g}}'$. *Then there exists a sequence of solutions* $\hat{\theta}_n$ *of* (5.1), *and there exist polynomials* p_j *in* q *variables* $(1 \le j \le r-2)$ *such that*

$$\sup_{B \in \mathcal{B}_{a,c}} \left| \mathbf{P}_\nu^{\theta_0}\{\sqrt{n}(\hat{\theta}_n - \theta_0) \in B\} - \int_B \left\{ \phi_{V(\mathbf{g})}(y) + \sum_{j=1}^{r-2} n^{-j/2} \phi_{j,V,\mathbf{g}}(y) \right\} dy \right|$$
$$= o\big(n^{-(r-2)/2}\big),$$

where $\phi_{j,V,\mathbf{g}} = \tilde{\pi}_{j,\mathbf{g}}(-D)\phi_V$ *and* $\tilde{\pi}_{j,\mathbf{g}}(y)$ *is a polynomial in* $y (\in \mathbf{R}^p)$ *whose coefficients are smooth functions of the partial derivatives of* $\lambda(\alpha)$ *at* $\alpha = 0$ *up to order* $j+2$, *and those of* $\nu_\alpha \mathbf{Q}_\alpha h_1$ *at* $\alpha = 0$ *up to order* j *together with those of* $\mathbf{g}$ *at* μ *up to order* $j+1$.

The application of Theorem 7 to third-order efficiency for the MLE and third-order efficient approximate solution of the likelihood equation follows directly from [28].

6. Examples. From a theoretical point of view, Theorems 5–7 are adequate for state space model estimation problems in providing assurance of the existence of efficient estimators, characterizing them as solutions of likelihood equations and prescribing their asymptotic behavior. In practice, however, one must still contend with certain statistical and numerical difficulties, such as implementation of the maximum likelihood estimator. In this section we apply our results to study some examples which include Markov switching models ARMA models, (G)ARCH models and SV models. For simplicity, in these examples we consider only specific structure of normal error assumption in most cases. Although strong consistency and asymptotic normality of the MLE in ARMA and GARCH(p,q) have been known in the literature, we provide alternative proofs in the framework of state space models. Furthermore, we can apply Theorem 7 to have Edgeworth expansion for the MLE. To the best of our knowledge, the asymptotic normality of the MLE in the AR(1)/ARCH(1) model, considered in Section 6.3, seems to be new. The results of asymptotic properties for the MLE in stochastic volatility models not only provide theoretical justification, but also give some insight into the structure of the likelihood function, which can be used for further study.

6.1. *Markov switching models.* We start with a simple real-valued fourth-order autoregression around one of two constants, μ_1 or μ_2:

$$\xi_n - \mu_{X_n} = \sum_{k=1}^{4} \varphi_k(\xi_{n-k} - \mu_{X_{n-k}}) + \varepsilon_n, \tag{6.1}$$

where $\varepsilon_n \sim N(0,\sigma^2)$, and $\{X_n, n \geq 0\}$ is a two-state Markov chain. This model was studied by Hamilton [33] in order to analyze the behavior of U.S. real GNP. To apply our theory in the form of (6.1), we consider a simple case of order 1 in (6.1). In this case, the likelihood function for given $X_n = x_n$, $n \geq 0$, is

$$f(\xi_n|x_n;\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-[(\xi_n - \mu_{x_n}) - \varphi_1(\xi_{n-1} - \mu_{x_{n-1}})]^2/(2\sigma^2)). \tag{6.2}$$

Denote by $[p_{xy}]_{x,y=1,2}$ the transition probability of the underlying Markov chain $\{X_n, n \geq 0\}$ and let $\theta = (p_{11}, p_{21}, \varphi_1, \mu_1, \mu_2, \sigma^2)$ be the unknown parameter. Assume that $|\varphi_1| < 1$, and that there exists a constant $c > 0$ such that $\sigma^2 > c$. Moreover, we assume that $\mu_1 \neq \mu_2$ such that the identifiability condition C6 holds. Since the state space of X_n is finite, we consider $0 < p_{xy} < 1$ for all $x, y = 1, 2$, and let $w(x) = |x| + 1$ such that the condition C1 holds. Under the normal distribution assumption, it is easy to see that conditions C2–C4 and C7–C9 are satisfied in this model. To check that C5 holds note that condition C5 reduces to

$$\sup_{x\in\mathcal{X}} E_x^{\theta^0}\left(\left[\sup_{|\theta-\theta^0|<\delta}\max_{y,z\in\mathcal{X}}\frac{f(\xi_0;\theta|y)f(\xi_1;\theta|y,\xi_0)}{f(\xi_0;\theta|z)f(\xi_1;\theta|z,\xi_0)}\right]^2\right) < \infty. \tag{6.3}$$

Since the maximum over x, y and z is applied to a finite set $\mathcal{X}$, and f defined in (6.1) is a normal density, it is easy to check that (6.3) is satisfied.

When $\xi_n = X_n$ as in (6.1), that is, $\mu_1 = \mu_2 = \mu$ are given, this reduces to the classical autoregressive model with unknown parameters $\theta = (\varphi_1, \dots, \varphi_4, \sigma^2)$. The Fisher information matrix is then given by

$$\mathbf{I}(\theta) = \begin{pmatrix} \sigma^{-2}\Gamma & 0 \\ 0 & 2(\sigma^4)^{-1} \end{pmatrix}, \tag{6.4}$$

where $\Gamma = (\gamma_{i-j})_{4\times 4}$ for $1 \le i, j \le 4$ with $\gamma_k = EX_nX_{n+k}$. A simple calculation shows that (5.9) reduces to (6.4) in this case. When $\varphi_k = 0$ as in (6.1), this is the hidden Markov model with normal mixture distributions considered in Example 1 of [7].

6.2. *ARMA models.* We start with a univariate Gaussian causal ARMA(p,q) model which can be written as a state space model by defining $r = \max\{p, q+1\}$,

$$\begin{aligned} \xi_n - \mu &= \alpha_1(\xi_{n-1} - \mu) + \alpha_2(\xi_{n-2} - \mu) + \cdots + \alpha_r(\xi_{n-r} - \mu) \\ &\quad + \varepsilon_n + \beta_1\varepsilon_{n-1} + \beta_2\varepsilon_{n-2} + \cdots + \beta_{r-1}\varepsilon_{n-r+1}, \end{aligned} \tag{6.5}$$

where $\alpha_j = 0$ for $j > p$ and $\beta_j = 0$ for $j > q$. Furthermore, we assume ε_n are i.i.d. random variables with distribution $N(0, \sigma^2)$. Asymptotic properties of the MLE in the ARMA model can be found in [35] and [53]. A general treatment of the MLE in the Gaussian ARMAX model can be found in Chapter 7 of [11].

By using the same idea as that in [34], we consider the following state space representation of (6.5):

$$X_{n+1} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{r-1} & \alpha_r \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} X_n + \begin{bmatrix} \varepsilon_{n+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6.6}$$

and

$$\xi_n = \mu + [1\ \beta_1\ \beta_2 \cdots \beta_{r-1}]X_n. \tag{6.7}$$

Assume that the roots of $1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p = 0$ lie outside the unit circle. It is easy to see that $\{X_n, n \ge 0\}$ forms a w-uniformly ergodic Markov chain with $w(x) = \|x\|^2$ (cf. Theorem 16.5.1 in [46]). And ξ_n are conditionally independent given $\{X_n, n \ge 0\}$. Since the verification of the weighted mean contraction property and the weighted moment assumption is the same as those in Remark 2(b), it will not be repeated here. This implies that condition C1 holds. The assumption $\varepsilon_n \sim N(0, \sigma^2)$ also implies that conditions C2–C5, C2′–C5′ and C7–C9 are satisfied in model (6.5). Since the verification is straightforward, we do not report it here. Suppose the conditional distribution of ξ_n given $X_0, \dots, X_n$ is of the form F_{X_{n-1},X_n} from (6.7). The Cramér conditions (4.16) and (4.17) hold for $Z_j^{(\nu)} := D^\nu \log p_1(\xi_0, \xi_1; \theta_0)$, since the conditional density of ξ_n given $\{x_n, n \ge 0\}$ is $N(0, \sigma^2)$ and

$$\limsup_{|\theta|\to 0} \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} e^{i\theta\xi}\, dF_{x,\alpha x+z}(\xi) \right\} \varphi(z)\, dz\, \pi(dx) \right| < 1, \tag{6.8}$$

where $\varphi(\cdot)$ is the normal density function of ε_1, and π is the stationary distribution of $\{X_n\}$. The identification issue in C6 can be found in Chapter 9 of [11] or Chapter 13 of [34].

6.3. *(G)ARCH models.* In this subsection we study two specific (G)ARCH models. To start with, we consider the AR(1)/ARCH(1) model

$$X_n = \beta_0 + \beta_1 X_{n-1} + \sqrt{\alpha_0 + \alpha_1 X_{n-1}^2}\,\varepsilon_n, \tag{6.9}$$

where α_i, β_i are unknown parameters for $i = 0, 1$ with $\alpha_0 > 0$, $0 < \alpha_1 < 1$, $3\alpha_1^2 < 1$ and $0 < \beta_1 < 1$. Here ε_n are i.i.d. random variables with the standard normal distribution. Note that in (6.9) $\mathbf{X} = (X_n)$ is defined as the autoregressive scheme AR(1) with ARCH(1) noise $(\sqrt{\alpha_0 + \alpha_1 X_{n-1}^2}\,\varepsilon_n)_{n\ge 1}$. When $\beta_0 = \beta_1 = 0$, this is the classical ARCH(1) model first considered by Engle [19].

Model (6.9) is conditionally Gaussian, and therefore the likelihood function of the parameter $\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1)$ for given observations $\mathbf{x} = (x_0 = 0, x_1, \ldots, x_n)$ from (6.9) is

$$\begin{aligned} l(\mathbf{x};\theta) = (2\pi)^{-n/2} \prod_{k=1}^{n} (\alpha_0 + \alpha_1 x_{k-1}^2)^{-1/2} \\ \times \exp\left\{ -\frac{1}{2} \sum_{k=1}^{n} \frac{(x_k - \beta_0 - \beta_1 x_{k-1})^2}{\alpha_0 + \alpha_1 x_{k-1}^2} \right\}. \end{aligned} \tag{6.10}$$

Assume $\beta_0 = 0$ and α_0, α_1 are given. The maximum likelihood estimator $\hat\beta_1$ of β_1 is the root of the equation $\partial l(\mathbf{x};\theta)/\partial\beta_1 = 0$. In view of (6.9) and (6.10), we obtain

$$\begin{aligned} \hat\beta_1 &= \frac{\sum_{k=1}^n (x_k - \beta_0) x_{k-1}/(\alpha_0 + \alpha_1 x_{k-1}^2)}{\sum_{k=1}^n x_{k-1}^2/(\alpha_0 + \alpha_1 x_{k-1}^2)} \\ &= \beta_1 + \frac{\sum_{k=1}^n x_{k-1}\varepsilon_k/\sqrt{\alpha_0 + \alpha_1 x_{k-1}^2}}{\sum_{k=1}^n x_{k-1}^2/(\alpha_0 + \alpha_1 x_{k-1}^2)}. \end{aligned} \tag{6.11}$$

Meyn and Tweedie [46], pages 380 and 383, establish w-uniform ergodicity [with $w(x) = |x| + 1$] of the AR(1) model $X_n = \beta_0 + \beta_1 X_{n-1} + \varepsilon_n$ by proving that a *drift condition* is satisfied, where $|\beta_1| < 1$ and the ε_n are i.i.d. random variables, with $E|\varepsilon_n| < \infty$, whose common density function q with respect to Lebesgue measure is positive everywhere. The strongly nonlattice condition holds as that in model (6.5). By using an argument similar to Theorem 1 of [45], we have the asymptotic identifiability of the likelihood function (6.10). Letting $\xi_n = X_n$, and using an argument similar to that in Remark 2(b), condition C1 holds. The verification of conditions C2–C9 and C2′–C5′ is straightforward and tedious, and is thus omitted. By Theorems 5–7, we have the strong consistency, asymptotic normality and Edgeworth expansion of the MLE $\hat{\beta}_1$. The asymptotic properties of the MLE of β_0, α_0 and α_1 can be verified in a similar way.

Next, we consider the GARCH(p,q) model of (1.1) in Example 1. It is known that the necessary and sufficient condition for (1.1) defining a unique strictly stationary process $\{Y_n, n \geq 0\}$ with $EY_n^2 < \infty$ is

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{6.12}$$

We assume (6.12) holds.

Similar to the estimation for ARMA models, the most frequently used estimators for GARCH models are those derived from a (conditional) Gaussian likelihood function (cf. [20]). Without the normal assumption of ε_n in (1.1), and imposing the moment condition $E(\varepsilon_1^4) < \infty$, Hall and Yao [32] established the asymptotic normality of the conditional maximum likelihood estimator in GARCH(p,q). They also established asymptotic results when the case of the error distribution is heavy-tailed. Earlier in the literature, when $p = q = 1$, Lee and Hansen [43] and Lumsdaine [45] proved, under some regularity conditions, the consistency and asymptotic normality for the quasi-maximum likelihood estimator in the GARCH(1, 1) model.

By using the state space representation (1.2) and (1.3), it is known (cf. Theorem 3.2 of [1]) that the Markov chain $\{X_n, n \geq 0\}$ defined in (1.3) is stationary if and only if the top Lyapunov exponent γ of A_n is strictly negative. It is easy to see that $\{X_n, n \geq 0\}$ is an aperiodic, irreducible and w-uniformly [with $w(x) = \|x\|^2$] ergodic Markov chain. Furthermore, we assume ε_n are i.i.d. random variables with distribution $N(0, \sigma^2)$. An argument similar to that in Remark 2(b) leads to condition C1 holding. The normal error assumption also implies that conditions C2–C5, C2′–C5′ and C7–C9 are satisfied in model (1.3). When $p = q = 1$, Theorem 1 of [45] proves the asymptotic identifiability of the likelihood function.

6.4. *Stochastic volatility models.* Consider the stochastic volatility model (1.4)–(1.8). To check that condition C1 holds, we note that $w(x) = |x| + 1$ in the AR(1) model $X_n = \alpha X_{n-1} + \eta_n$ by proving that a *drift condition* is satisfied, where $|\alpha| < 1$ and the η_n are i.i.d. random variables, with $E|\eta_1| < \infty$, whose common density function q with respect to Lebesgue measure is positive everywhere.

Since $\varepsilon_n \sim N(0,1)$, $\zeta_n = \log \varepsilon^2$, $\eta_n \sim N(0,\sigma_\eta^2)$, and ζ_n and η_n are mutually independent, an argument similar to that in Remark 2(b) leads to the result that the rest of condition C1 holds. Conditions C2–C5, C2′–C5′ and C7–C9 are also satisfied in model (1.5) and (1.6) (cf. pages 22–23 of [50]). Denote $\xi_n := \log Y_n^2$. Note that the conditional density of X_n exists, and this implies that the conditional distribution of ξ_n given $X_0, \dots, X_n$ is of the form F_{X_{n-1},X_n} such that

$$\limsup_{|t|\to 0} \left| \int_{\mathcal{X}} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} e^{its}\, dF_{x,\alpha x+z}(s) \right\} \varphi(z)\, dz\, \pi(dx) \right| < 1, \tag{6.13}$$

where $\varphi(\cdot)$ is the normal density function of ζ_1 and π is the stationary distribution of $\{X_n\}$. Let $S_n = \sum_{i=1}^n \xi_i$, $S_0 = 0$. Then $\{(X_n, S_n), n \geq 0\}$ is strongly nonlattice. To check the identification condition C6, the reader is referred to Chapter 13 of [34] and Section 2.4.3 of [29].

Next, we assume that $\varepsilon_n \sim N(0,1)$, $\zeta_n = \log \varepsilon_n^2$ and η_n is a sequence of i.i.d. double exponential(1) random variables. Furthermore, we assume ζ_n and η_n are mutually independent. By using an argument similar to that in Remark 2(b), condition C1 holds. Simple calculations also lead conditions C2–C5, C2′–C5′ and C7–C9 to hold in this case. Under the assumption that the conditional distribution of ξ_n given $X_0, \dots, X_n$ is of the form F_{X_{n-1},X_n} such that (6.13) holds, $\{(X_n, S_n), n \geq 0\}$ is strongly nonlattice.

Without the normal assumption, quasi-maximum likelihood (QML) estimators of the parameters are obtained by treating ζ_n and η_n as though they were normal and maximizing the prediction error decomposition form of the likelihood obtained via the Kalman filter or implied volatility. That is, we assume that ζ_n is a sequence of independent and identically distributed $N(0,\sigma_\zeta^2)$ random variables. For given observations $\mathbf{y} = (\log y_1^2, \dots, \log y_n^2)$ from (1.5) and (1.6), the likelihood function of the parameter $\theta = (\alpha, \sigma_\eta^2, \sigma_\zeta^2)$ is

$$\begin{aligned} l(\mathbf{y};\theta) = {} & \int_{x_0\in\mathcal{X}} \cdots \int_{x_n\in\mathcal{X}} \pi(x_0)(2\pi\sigma_\zeta^2)^{-n/2} \\ & \times \prod_{k=1}^n p(x_{k-1},x_k) \\ & \times \exp\left\{ -\frac{1}{2} \sum_{k=1}^n \frac{(\log y_k^2 - \omega - x_k)^2}{\sigma_\zeta^2} \right\} dx_0\, dx_1 \cdots dx_n, \end{aligned} \tag{6.14}$$

where $p(x_{k-1}, x_k)$ is defined in (1.7). By using the results of [16], Harvey, Ruiz and Shephard [36] showed that the quasi-maximum likelihood estimators are asymptotically normal under some regularity conditions. Further study of the MLE in stochastic volatility models will be published in a separate paper.

7. Proofs of Lemmas 3–6. For convenience of notation, denote $\{Z_n, n \geq 0\} := \{((X_n, \xi_n), M_n), n \geq 0\}$ as the Markov chain induced by the Markovian iterated random functions system (2.4)–(2.7) on the state space $(\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$. In the proof of Lemma 3, we omit θ in $\mathbf{P}_\theta(\cdot)$ for simplicity.

PROOF OF LEMMA 3. We consider only the cases of $\mathbf{P}(\xi_1)$, since the cases of $\mathbf{P}(\xi_0)$ and $\mathbf{P}(\xi_j)$, for $j = 2, \dots, n$, are a straightforward consequence. For any two elements $h_1, h_2 \in \mathbf{M}$, and two fixed elements $s_0, s_1 \in \mathbf{R}^d$, by (5.8) we have

$$
\begin{aligned}
d&\big(\mathbf{P}(s_1)h_1, \mathbf{P}(s_1)h_2\big) \\
&= \sup_{x_0 \in \mathcal{X}} \biggl| \int p_\theta(x_0, x_1) f(s_1; \theta | x_1, s_0) h_1(x_1) m(dx_1) \\
&\qquad\qquad - \int p_\theta(x_0, x_1) f(s_1; \theta | x_1, s_0) h_2(x_1) m(dx_1) \biggr| \\
&\leq d(h_1, h_2) \sup_{x_0 \in \mathcal{X}} \int p_\theta(x_0, x_1) f(s_1; \theta | x_1, s_0) m(dx_1) \\
&\leq C \biggl(\sup_{x_0 \in \mathcal{X}} \int p_\theta(x_0, x_1) m(dx_1) \biggr) d(h_1, h_2),
\end{aligned}
$$

where $0 < C = \sup_{x_1 \in \mathcal{X}} f(s_1; \theta | x_1, s_0) < \infty$ by assumption C1 is a constant. Note that $\sup_{x_0 \in \mathcal{X}} \int p_\theta(x_0, x_1) m(dx_1) = 1$. The equality holds only if $h_1 = h_2$ m-almost surely. This proves the Lipschitz continuous condition in the second argument.

Note that C1 implies Assumption K1 holds. Recall that $M_n = \mathbf{P}(\xi_n) \circ \cdots \circ \mathbf{P}(\xi_1) \circ \mathbf{P}(\xi_0)$ in (5.6). To prove the weighted mean contraction property K2, we observe that, for $p \geq 1$,

$$
\begin{aligned}
\sup_{x_0, s_0} & \mathbf{E}_{(x_0, s_0)} \biggl\{ \log \biggl(L_p \frac{w(X_p, \xi_p)}{w(x_0, s_0)} \biggr) \biggr\} \\
&= \sup_{x_0, s_0} \mathbf{E}_{(x_0, s_0)} \biggl\{ \log \biggl(\sup_{h_1 \neq h_2} \frac{d(M_p h_1, M_p h_2)}{d(h_1, h_2)} \frac{w(X_p, \xi_p)}{w(x_0, s_0)} \biggr) \biggr\} \\
&< \sup_{x_0, s_0} \mathbf{E}_{(x_0, s_0)} \biggl\{ \log \biggl(\biggl[\sup_{x_0 \in \mathcal{X}} \int p_\theta(x_0, x_1) f(\xi_1; \theta | x_1, s_0) m(dx_1) \biggr]^p \\
&\qquad\qquad \times \frac{w(X_p, \xi_p)}{w(x_0, s_0)} \biggr) \biggr\} \\
&< 0.
\end{aligned}
\tag{7.1}
$$

The last inequality follows from (5.2) in condition C1.

To verify that Assumption K3 holds, as m is σ-finite, we have $\mathcal{X} = \bigcup_{n=1}^{\infty} \mathcal{X}_n$, where the $\mathcal{X}_n$ are pairwise disjoint and $0 < m(\mathcal{X}_n) < \infty$. Set

$$h(x) = \sum_{n=1}^{\infty} \frac{I_{\mathcal{X}_n}(x)}{2^n m(\mathcal{X}_n)}. \tag{7.2}$$

It is easy to see that $\int_{x \in \mathcal{X}} h(x)m(dx) = 1$ and, hence, belongs to $\mathbf{M}$. Observe that

$$\begin{aligned} &\mathbf{E}d^2(\mathbf{P}(\xi_j)h, h) \\ &\quad = \mathbf{E} \sup_{x_{j-1} \in \mathcal{X}} \biggl| \int p_\theta(x_{j-1}, x_j) \\ &\qquad\qquad \times f(\xi_j; \theta | x_j, \xi_{j-1}) h(x_j) m(dx_j) - h(x_{j-1}) \biggr|. \end{aligned} \tag{7.3}$$

By definition of $h(x)$ in (7.2), it is piecewise constant, and $p_\theta(x_{j-1}, x_j) f(\xi_j; \varphi_{x_j}(\theta)|\xi_{j-1})$ is a probability density function integrable over the subset $\mathcal{X}_n$. These imply (7.3) is finite.

Finally, we observe

$$\begin{aligned} &\sup_{x_0, s_0} \mathbf{E}_{(x_0, s_0)} \left\{ L_1 \frac{w(X_1, \xi_1)}{w(x_0, s_0)} \right\} \\ &\quad = \sup_{x_0, s_0} \mathbf{E}_{(x_0, s_0)} \left\{ \sup_{h_1 \neq h_2} \frac{d(\mathbf{P}(\xi_1)h_1, \mathbf{P}(\xi_1)h_2)}{d(h_1, h_2)} \frac{w(X_1, \xi_1)}{w(x_0, s_0)} \right\} \\ &\quad < \sup_{x_0, s_0} \mathbf{E}_{(x_0, s_0)} \left\{ \sup_{x_0 \in \mathcal{X}} \int p_\theta(x_0, x_1) f(\xi_1; \theta | x_1, s_0) m(dx_1) \frac{w(X_1, \xi_1)}{w(x_0, s_0)} \right\} < \infty. \end{aligned}$$

The last inequality follows from (5.3) in condition C1.

Note that C5 implies the exponential moment condition of g. Hence, the proof is complete. $\square$

In the proof of Lemma 4 we omit θ for simplicity.

PROOF OF LEMMA 4. We first prove that $\{Z_n, n \geq 0\}$ is Harris recurrent. Note that the transition probability kernel of the Markov chain $\{(X_n, \xi_n), n \geq 0\}$, defined in (2.1) and (2.2), has a probability density with respect to $m \times Q$. And the iterated random functions system, defined in (2.4)–(2.7), also has a probability density with respect to Q. By making use the definition (3.2), there exists a measurable function $g : (\mathcal{X} \times \mathbf{R}^d \times \mathbf{M}) \times (\mathcal{X} \times \mathbf{R}^d \times \mathbf{M}) \to [0, \infty)$ such that

$$\mathbf{P}(z, dz') = g(z, z')(m \times Q \times Q)(dz'), \tag{7.4}$$

where $\int_{(\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}} g(z, z')(m \times Q \times Q)(dz') = 1$ for all $z \in (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$. For simplicity of notation, we let $\Lambda(\cdot) := (m \times Q \times Q)(\cdot)$ in the proof. For given

$n > 1$, let $\mathbf{P}^n(z, \cdot) := \mathbf{P}_z(Z_n \in \cdot)$ for $z \in (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$. For $A \in \mathcal{B}(\mathcal{X} \times \mathbf{R}^d)$ and $B \in \mathcal{B}(\mathbf{M})$, define

$$\Lambda^n(A \times B) := \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} \mathbf{P}_{z'}\{Z_n \in A \times B\}\Lambda(dz').$$

Then for all $A \in \mathcal{B}(\mathcal{X} \times \mathbf{R}^d)$ and $B \in \mathcal{B}(\mathbf{M})$,

$$\begin{aligned}\mathbf{P}^{n+1}(z, A \times B) &= \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} \mathbf{P}^n(z', A \times B) g(z, z')\Lambda(dz') \\ &= \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} \mathbf{P}_{z'}\{Z_n \in A \times B\} g(z, z')\Lambda(dz').\end{aligned}$$

It is easy to see that, for given any $n > 1$, the family $(\mathbf{P}^{n+1}(z, \cdot))_{z\in(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}}$ is absolutely continuous with respect to Λ^n. Therefore, by the Radon–Nikodym theorem, $\mathbf{P}^n$ has a probability density with respect to Λ^n for all $n \geq 1$. Let g_n be such that

$$\mathbf{P}^{n+1}(z, dz') = g_n(z, z')\Lambda^n(dz'), \qquad z \in (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}, \tag{7.5}$$

where $\int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} g_n(z, z')\Lambda^n(dz') = 1$ for all $z \in (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$. Note that $g_1 = g$. It is easy to check that all Λ^n are absolutely continuous with respect to Π.

Denote B^c as the complement of B. Since $\Pi(((\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M})^c) = 0$, also $\Lambda(((\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M})^c) = 0$. Recall g is defined in (7.4). It is obvious from the previous considerations that we can choose $\delta > 0$ sufficiently small such that

$$\begin{aligned}\int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} & \mathbb{1}_{\{g_2\geq\delta\}}(z_1, z_2) \\ & \times \mathbb{1}_{\{g\geq\delta\}}(z_2, z_3)\Lambda(dz_3)\Lambda^2(dz_2)\Pi(dz_1) > 0.\end{aligned}$$

Hence, by Lemma 4.3 of [48], there exist a Π-positive set $\Gamma_1 \subset (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$ and a Λ-positive set $\Gamma_2 \subset (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$ such that

$$\alpha := \inf_{z_1\in\Gamma_1, z_3\in\Gamma_2} \Lambda^2\{z_2 \in (\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M} : g_2(z_1, z_2) \geq \delta, g(z_2, z_3) \geq \delta\} > 0.$$

A combination of the above result with (7.4) and (7.5) implies

$$\begin{aligned}\mathbf{P}^3(z_1, A \times B) &= \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} \mathbf{P}(z_2, A \times B)\mathbf{P}^2(z_1, dz_2) \\ &\geq \int_{(\mathcal{X}\times\mathbf{R}^d)\times\mathbf{M}} g_2(z_1, z_2) \int_{(A\times B)\cap\Gamma_2} g(z_2, z_3)\Lambda(dz_3)\Lambda^2(dz_2) \\ &\geq \alpha\delta^2\Lambda\big((A \times B) \cap \Gamma_2\big)\end{aligned} \tag{7.6}$$

for all $z_1 \in \Gamma_1$ and $A \times B \in \mathcal{B}((\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M})$. Therefore, we obtain an absorbing set such that Γ_1 is a regeneration set for $\{Z_n, n \geq 0\}$ on $(\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$, that is,

Γ_1 is recurrent and satisfies a minorization condition, namely, (7.6). This proves the Harris recurrence of $\{Z_n, n \geq 0\}$ on $(\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$. Since $\{Z_n, n \geq 0\}$ possesses a stationary distribution, it is clearly positive Harris recurrent.

Next, we give the proof of aperiodicity. If $\{Z_n, n \geq 0\}$ were q-periodic with cyclic classes $\Gamma_1, \dots, \Gamma_q$, say, then the q-skeleton $(Z_{nq})_{n\geq 0}$ would have stationary distributions $\frac{\Pi(\cdot \cap \Gamma_k)}{\Pi(\Gamma_k)}$ for $k = 1, \dots, q$. On the other hand, Z_{qn} is aperiodic by definition, and M_{nq} is also a Markovian iterated random functions system of Lipschitz maps, satisfying condition C1, and thus possesses only one stationary distribution. Consequently, $q = 1$ and $\{Z_n, n \geq 0\}$ is aperiodic. Since the Markov chain $\{((X_n, \xi_n), M_n), n \geq 0\}$ has a probability density with respect to Λ, it is obviously Λ-irreducible. The proof is complete. $\square$

PROOF OF LEMMA 5. In order to define the Fisher information (5.9), we need to verify that there exists a $\delta > 0$, such that $\partial \log \|\mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|/\partial\theta \in L_2(\mathbf{P}_\Pi^\theta)$ for $\theta \in N_\delta(\theta_0)$, a δ-neighborhood of θ_0. That is, we need to show

$$\mathbf{E}_\Pi^\theta \left(\frac{\partial \log \|\mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|}{\partial\theta} \right)^2 < \infty, \tag{7.7}$$

for $\theta \in N_\delta(\theta_0)$.

It is easy to see that C5 implies that

$$\sup_{x\in\mathcal{X}} E_x^\theta \left(\frac{\partial \log \int_{y\in\mathcal{X}} \pi(x)p(x, y)f(\xi_0; \theta|x)f(\xi_1; \theta|y, \xi_0)m(dy)}{\partial\theta} \right)^2 < \infty$$

for $\theta \in N_\delta(\theta_0)$. And this leads to

$$\sup_{x\in\mathcal{X}} \mathbf{E}_x^\theta \left(\frac{\partial \log \int_{y\in\mathcal{X}} \pi(x)p(x, y)f(\xi_0; \theta|x)f(\xi_1; \theta|y, \xi_0)m(dy)}{\partial\theta} \right)^2 < \infty \tag{7.8}$$

for $\theta \in N_\delta(\theta_0)$, where $\mathbf{E}_x^\theta$ is the expectation under $\mathbf{P}^\theta(\cdot, \cdot)$.

Finally, (7.8) implies (7.7) and we have the proof. $\square$

PROOF OF LEMMA 6. For each $j = 1, \dots, q$,

$$\begin{aligned}
\frac{1}{\sqrt{n}} l_j'(\theta_0) &= \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta_j} \log \|\mathbf{P}_\theta(\xi_n) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\| \Big|_{\theta=\theta_0} \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^n \left(\frac{\partial}{\partial\theta_j} \log \frac{\|\mathbf{P}_\theta(\xi_k) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|}{\|\mathbf{P}_\theta(\xi_{k-1}) \circ \cdots \circ \mathbf{P}_\theta(\xi_1) \circ \mathbf{P}_\theta(\xi_0)\pi\|} \Big|_{\theta=\theta_0} \right) \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\partial}{\partial\theta_j} g(M_{k-1}, M_k) \Big|_{\theta=\theta_0}.
\end{aligned}$$

Now, for each $h \in \mathbf{M}$, $\alpha = (\alpha_1, \dots, \alpha_q) \in C^q$, and a $(\mathcal{X} \times \mathbf{R}^d) \times \mathbf{M}$ measurable function φ with $\|\varphi\|_{wh} < \infty$, define

$$
\begin{aligned}
&(\mathbf{T}_1(\alpha)\varphi)\big((x,s),h\big)\\
&\quad = \mathbf{E}^{\theta_0}_{(x,s)}\bigg\{\exp\bigg((\alpha_1,\dots,\alpha_q)'\bigg(\frac{\partial}{\partial\theta_1}\log\|\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)h\|\bigg|_{\theta=\theta_0},\dots,\\
&\qquad\qquad \frac{\partial}{\partial\theta_q}\log\|\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)h\|\bigg|_{\theta=\theta_0}\bigg)\bigg)\\
&\qquad\qquad \times \varphi\big((X_1,\xi_1),\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)h(x)\big)\bigg\}.
\end{aligned}
\tag{7.9}
$$

By using an argument similar to that of Lemma 2, we have, for sufficiently small $|\alpha|$, $\mathbf{T}_1(\alpha)$ is a bounded and analytic operator. Let $\lambda^{\theta_0}_{\mathbf{T}_1}(\alpha)$ be the eigenvalue of $\mathbf{T}_1(\alpha)$ corresponding to a one-dimensional eigenspace. Define γ_j as that in Lemma 2(v). By conditions C1–C5 and Lemma 4, it is easy to see that

$$
\gamma_j = \frac{\partial}{\partial\alpha_j}\lambda^{\theta_0}_{\mathbf{T}_1}(\alpha)\bigg|_{\alpha=0} = \mathbf{E}^{\theta_0}_\Pi\bigg(\frac{\partial}{\partial\theta_j}\log\|\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)\pi\|\bigg|_{\theta=\theta_0}\bigg) = 0. \tag{7.10}
$$

By Corollary 1, we have

$$
\frac{1}{\sqrt{n}}(l'_j(\theta_0))_{j=1,\dots,q} \longrightarrow N\big(0,\boldsymbol{\Sigma}(\theta_0)\big) \qquad \text{in distribution,} \tag{7.11}
$$

where the variance–covariance matrix

$$
\boldsymbol{\Sigma}(\theta_0) = (\Sigma_{ij}(\theta_0)) = \bigg(\frac{\partial^2\lambda^{\theta_0}_{\mathbf{T}_1}(\alpha)}{\partial\alpha_i\,\partial\alpha_j}\bigg|_{\alpha=0}\bigg)_{i,j=1,\dots,q}. \tag{7.12}
$$

In the following, we will verify that the variance–covariance matrix $\boldsymbol{\Sigma}(\theta_0)$ defined as (7.12) is the Fisher information matrix $\mathbf{I}(\theta_0)$. By Lemma 2 and Corollary 1, we have

$$
\begin{aligned}
&\mathbf{E}^{\theta_0}_\Pi\bigg(\bigg(\frac{\partial}{\partial\theta_j}\log\|M_n\pi\|\bigg|_{\theta=\theta_0}\bigg)\bigg(\frac{\partial}{\partial\theta_k}\log\|M_n\pi\|\bigg|_{\theta=\theta_0}\bigg)\bigg) - n\frac{\partial^2}{\partial\alpha_j\,\partial\alpha_k}\lambda^{\theta_0}_{\mathbf{T}_1}(\alpha)\bigg|_{\alpha=0}\\
&\qquad \longrightarrow 0
\end{aligned}
$$

as $n \to \infty$. Therefore,

$$
\begin{aligned}
\Sigma_{jk}(\theta_0) &= \frac{\partial^2}{\partial\alpha_j\,\partial\alpha_k}\lambda^{\theta_0}_{\mathbf{T}_1}(\alpha)\bigg|_{\alpha=0}\\
&= \lim_{n\to\infty}\frac{1}{n}\mathbf{E}^{\theta_0}_\Pi\bigg(\frac{\partial}{\partial\theta_j}\log\|M_n\pi\|\bigg|_{\theta=\theta_0}\bigg)\bigg(\frac{\partial}{\partial\theta_k}\log\|M_n\pi\|\bigg|_{\theta=\theta_0}\bigg)\\
&= \lim_{n\to\infty} -\frac{1}{n}\mathbf{E}^{\theta_0}_\Pi\bigg(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log\|M_n\pi\|\bigg|_{\theta=\theta_0}\bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= -\mathbf{E}_{\Pi}^{\theta_0}\left(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log\|\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)\pi\|\Big|_{\theta=\theta_0}\right)\\
&= \mathbf{E}_{\Pi}^{\theta_0}\left(\frac{\partial}{\partial\theta_j}\log\|\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)\pi\|\Big|_{\theta=\theta_0}\right)\\
&\quad\times\left(\frac{\partial}{\partial\theta_k}\log\|\mathbf{P}_\theta(\xi_1)\circ\mathbf{P}_\theta(\xi_0)\pi\|\Big|_{\theta=\theta_0}\right)\\
&= I_{jk}(\theta_0). \qquad \square
\end{aligned}
$$

APPENDIX

Proofs of Lemma 1 and Theorem 2. In the following proofs we will use the same notation as in Sections 3 and 4 unless specified. Without loss of generality, in this section we consider the case $M_0 = \mathrm{Id}$, the identity, and the transition probability $\mathbf{P}$ of the Markov chain $\{(Y_n, M_n), n \geq 0\}$ depends on the initial state $Y_0 = y$ only. Denote it as $\mathbf{P}_y$, and let $\mathbf{E}_y$ be the corresponding expectation. To prove Lemma 1, we need the following lemma first.

LEMMA A.1. *Let $\{(Y_n, M_n), n \geq 0\}$ be the MIRFS of Lipschitz functions defined in (2.1) satisfying Assumption K. There exists $0 < \delta_0 < 1$ such that, for all $0 < \delta \leq \delta_0$, there exist $K > 0$, and $0 < \eta < 1$, so that*

$$
\sup_y \mathbf{E}_y\left\{\left(\frac{d(M_n^u, M_n^v)}{d(u,v)}\frac{w(Y_n)}{w(y)}\right)^\delta\right\} \leq K\eta^n, \qquad \text{for } n \in N \text{ and } u, v \in \mathbf{M}.
$$

PROOF. For given $0 < \delta < 1$, and $y \in \mathcal{Y}$, denote

$$
c_n(y) = \sup\left\{\mathbf{E}_y\left[\left(\frac{d(M_n^u, M_n^v)}{d(u,v)}\frac{w(Y_n)}{w(y)}\right)^\delta\right] : u, v \in \mathbf{M}\right\},
$$

and let $\eta_n = \sup\{c_n(y), y \in \mathcal{Y}\}$. Denote $u_m = M_m^u$ and $v_m = M_m^v$. Let $\mathcal{F}_m$ be the σ-algebra generated by $\{(Y_k, M_k), 0 \leq k \leq m\}$. Then

$$
\begin{aligned}
&\mathbf{E}_y\left\{\left(\frac{d(M_{n+m}^u, M_{n+m}^v)}{d(u,v)}\frac{w(Y_{n+m})}{w(y)}\right)^\delta\Big|\mathcal{F}_m\right\}\\
&\quad= \mathbf{E}_y\left\{\left(\frac{d(F_{n:m}(M_m^u), F_{n:m}(M_m^v))}{d(u,v)}\frac{w(Y_{n+m})}{w(y)}\right)^\delta\Big|\mathcal{F}_m\right\}\\
&\quad= \left(\frac{d(M_m^u, M_m^v)}{d(u,v)}\frac{w(Y_m)}{w(y)}\right)^\delta \mathbf{E}_y\left\{\left(\frac{d(F_{n:m}(u_m), F_{n:m}(v_m))}{d(u_m,v_m)}\frac{w(Y_{n+m})}{w(Y_m)}\right)^\delta\Big|\mathcal{F}_m\right\}\\
&\quad= \left(\frac{d(M_m^u, M_m^v)}{d(u,v)}\frac{w(Y_m)}{w(y)}\right)^\delta \mathbf{E}_{Y_m}\left\{\left(\frac{d(M_n^{u_m}, M_n^{v_m})}{d(u_m,v_m)}\frac{w(Y_{n+m})}{w(Y_m)}\right)^\delta\right\}\\
&\quad\leq \left(\frac{d(M_m^u, M_m^v)}{d(u,v)}\frac{w(Y_m)}{w(y)}\right)^\delta c_n(Y_m) \leq \eta_n\left(\frac{d(M_m^u, M_m^v)}{d(u,v)}\frac{w(Y_m)}{w(y)}\right)^\delta.
\end{aligned}
$$

This implies that

$$\mathbf{E}_y\left\{\left(\frac{d(M_{n+m}^u, M_{n+m}^v)}{d(u,v)}\frac{w(Y_{n+m})}{w(y)}\right)^\delta\right\} \\ \le \eta_n \mathbf{E}_y\left\{\left(\frac{d(M_m^u, M_m^v)}{d(u,v)}\frac{w(Y_m)}{w(y)}\right)^\delta\right\},$$

or $\eta_{n+m} \le \eta_n \eta_m$. Therefore,

$$\lim_{n\to\infty} \eta_n^{1/n} = \inf\{\eta_n^{1/n}, n \in N\}. \tag{A.1}$$

It is known by Assumption K2 that there exist $p \ge 1$ and $d > 0$ such that $\sup_y \mathbf{E}_y\{\log(\frac{d(M_p^u, M_p^v)}{d(u,v)}\frac{w(Y_p)}{w(y)})\} < -d < 0$. Along with $\sup_y \mathbf{E}_y\{\frac{w(Y_p)}{w(y)}\} < \infty$ by (4.2) and $\sup_y \mathbf{E}_y\{\frac{l(F_1)w(Y_1)}{w(y)}\} < \infty$ by Assumption K3, we have

$$\eta_p \le \sup_{y\in\mathcal{Y}} \mathbf{E}_y\left\{\left(l(F_1)^p \frac{w(Y_p)}{w(y)}\right)^\delta\right\} := \sup_{y\in\mathcal{Y}} \mathbf{E}_y\left\{\exp\left(\delta G_p + \delta \log \frac{w(Y_p)}{w(y)}\right)\right\} < \infty,$$

where $G_p = p \log l(F_1)$.

Since $e^y \le 1 + y + y^2 e^{|y|}/2$, we have, for $y \in \mathcal{Y}$, $u, v \in \mathbf{M}$,

$$\mathbf{E}_y\left\{\left(\frac{d(M_p^u, M_p^v)}{d(u,v)}\frac{w(Y_p)}{w(y)}\right)^\delta\right\} \\ \le 1 + \delta \mathbf{E}_y\left\{\log\left(\frac{d(M_p^u, M_p^v)}{d(u,v)}\frac{w(Y_p)}{w(y)}\right)\right\} \\ + \delta^2 \mathbf{E}_y\left\{\left(G_p + \log \frac{w(Y_p)}{w(y)}\right)^2 \exp\left(\delta G_p + \delta \log \frac{w(Y_p)}{w(y)}\right)\right\}.$$

For $u, v \in \mathbf{M}$, we have

$$\eta_p \le 1 - d\delta + \delta^2 \sup_{y\in\mathcal{Y}} \mathbf{E}_y\left\{\left(G_p + \log \frac{w(Y_p)}{w(y)}\right)^2 \exp\left(\delta G_p + \delta \log \frac{w(Y_p)}{w(y)}\right)\right\}.$$

Therefore, we can choose $\delta_0 > 0$ small enough so that $\eta_p < 1$. Along with (A.1), we obtain the proof. $\square$

PROOF OF LEMMA 1. For given $\varphi \in \mathcal{H}$, $y \in \mathcal{Y}$, and $u, v \in \mathbf{M}$, if $m \le n$, we have, for $0 < \delta \le \delta_0 < 1$,

$$\left|\mathbf{T}^n \varphi(y,u) - \mathbf{E}_y \varphi(Y_n, F_{n:m}(v))\right| / w(y) \\ = \left|\mathbf{E}_y \varphi(Y_n, M_n^u) - \mathbf{E}_y \varphi(Y_n, F_{n:m}(v))\right| / w(y) \\ \le \|\varphi\|_h \mathbf{E}_y\{d(M_n^u, F_{n:m}(v))^\delta w(Y_n)^\delta\} / w(y)$$

$$
\begin{aligned}
&\le \|\varphi\|_h \mathbf{E}_y\left\{\mathbf{E}_y\left[\left(d\big(F_{n:m}(M^u_{n-m}), F_{n:m}(v)\big)\frac{w(Y_n)}{w(Y_{n-m})}\right)^{\delta}\middle|\mathcal{F}_{n-m}\right]\frac{w(Y_{n-m})^{\delta}}{w(y)}\right\}\\
&\le \|\varphi\|_h \mathbf{E}_y\left\{\sup_{u,v\in\mathbf{M}}\mathbf{E}_{Y_{n-m}}\left[\left(d(M^u_m, M^v_m)\frac{w(Y_n)}{w(Y_{n-m})}\right)^{\delta}\right]\frac{w(Y_{n-m})^{\delta}}{w(y)}\right\}\\
&\le \|\varphi\|_h \mathbf{E}_y\left\{\sup_{u,v\in\mathbf{M}}\mathbf{E}_{Y_{n-m}}\left[\left(\frac{d(M^u_m, M^v_m)}{d(u,v)}\frac{w(Y_n)}{w(Y_{n-m})}\right)^{\delta}\right]\frac{w(Y_{n-m})}{w(y)}\right\}.
\end{aligned}
$$

Note that in the last inequality we use $d(u,v)\le 1$ and $w(y)\ge 1$ for all $y\in\mathcal{Y}$.

By making use of Lemma A.1, and $\sup_{y\in\mathcal{Y}} E_y[w(Y_1)/w(y)]<\infty$ in (4.2), there exist $K>0$ and $0<\eta<1$ such that

$$
\big|\mathbf{T}^n\varphi(y,u)-\mathbf{E}_y\varphi\big(Y_n, F_{n:m}(v)\big)\big|/w(y)\le\|\varphi\|_h K\eta^m\le\|\varphi\|_{wh}K\eta^m. \tag{A.2}
$$

Denote $h(y)=\mathbf{E}_y\varphi(Y_m, F_m(v))$. Then by assumption (4.1), there exist $\gamma>0$ and $0<\rho<1$ such that

$$
\begin{aligned}
&\big|\mathbf{E}_y\varphi\big(Y_n, F_{n:m}(v)\big)-\mathbf{E}_\Pi\varphi\big(Y_m, F_m(v)\big)\big|/w(y)\\
&\quad\le\big|\mathbf{E}_y\big\{\mathbf{E}_{Y_{n-m}}\varphi\big(Y_m, F_m(v)\big)\big\}-\mathbf{E}_\Pi\varphi\big(Y_m, F_m(v)\big)\big|/w(y)\\
&\quad\le\left|\mathbf{E}_y h(Y_{n-m})-\int h(y)\Pi(dy)\right|\Big/w(y)\\
&\quad\le\|\varphi\|_{wh}\gamma\rho^{n-m}.
\end{aligned} \tag{A.3}
$$

For given $m,k\in N$, by using Lemma A.1 again we have

$$
\begin{aligned}
&\big|\mathbf{E}_\Pi\varphi\big(Y_m, F_m(v)\big)-\mathbf{E}_\Pi\varphi\big(Y_{m+k}, F_{m+k}(v)\big)\big|/w(y)\\
&\quad\le\mathbf{E}_\Pi\big\{\big|\varphi\big(Y_{m+k}, F_{m+k:k}(v)\big)-\varphi\big(Y_{m+k}, F_{m+k:k}(M^v_k)\big)\big|\big\}/w(y)\\
&\quad\le\|\varphi\|_h\mathbf{E}_\Pi\left\{d\big(F_{m+k:k}(v), F_{m+k:k}(M^v_k)\big)^{\delta}\frac{w(Y_{m+k})^{\delta}}{w(y)}\right\}\\
&\quad\le\|\varphi\|_h\mathbf{E}_\Pi\left\{\sup_{u,v\in\mathbf{M}}\mathbf{E}_{Y_m}\left[\left(\frac{d(M^u_m, M^v_m)}{d(u,v)}\frac{w(Y_{m+k})}{w(Y_m)}\right)^{\delta}\right]\frac{w(Y_m)}{w(y)}\right\}\\
&\quad\le\|\varphi\|_{wh}K\eta^m.
\end{aligned}
$$

By making use of (A.2), (A.3) and the above inequality, we have that for any given $n\ge m$, $k\ge 0$, and for all $u,v\in\mathbf{M}$,

$$
\big|\mathbf{T}^n\varphi(y,u)-\mathbf{E}_\Pi\varphi\big(Y_{m+k}, F_{m+k}(v)\big)\big|/w(y)\le\|\varphi\|_{wh}(2K\eta^m+\gamma\rho^{n-m}).
$$

By setting $m=n/2$, we have that there exist $A>0$ and $0<r<1$ such that

$$
\|\mathbf{T}^n\varphi(y,u)-\mathbf{Q}\varphi(y,u)\|_w\le\|\varphi\|_{wh}Ar^n. \tag{A.4}
$$

On the other hand, for $u, v \in \mathbf{M}$,

$$
\begin{aligned}
&\frac{|(\mathbf{T}^n - \mathbf{Q})\varphi(y,u) - (\mathbf{T}^n - \mathbf{Q})\varphi(y,v)|}{(w(y)\,d(u,v))^\delta} \\
&\quad = \biggl|\mathbf{E}_y\varphi(Y_n, M_n^u) - \int \varphi(y,u)\Pi(dy\times du) \\
&\qquad\quad - \mathbf{E}_y\varphi(Y_n, M_n^v) + \int \varphi(y,v)\Pi(dy\times dv)\biggr| \\
&\qquad\quad \times \bigl[\bigl(w(y)\,d(u,v)\bigr)^{-\delta}\bigr]^{-1} \\
&\quad \le \frac{\mathbf{E}_y\{|\varphi(Y_n, M_n^u) - \varphi(Y_n, M_n^v)|\}}{(w(y)\,d(u,v))^\delta} \\
&\quad \le \|\varphi\|_h \sup_y \mathbf{E}_y\biggl\{\biggl(\frac{d(M_n^u, M_n^v)}{d(u,v)}\frac{w(Y_n)}{w(y)}\biggr)^\delta\biggr\} \\
&\quad \le \|\varphi\|_{wh} K\eta^n \qquad \text{by Lemma A.1.}
\end{aligned}
\tag{A.5}
$$

Denote $\rho_* = \min\{\eta, r\}$ and $\gamma_* = A + K$. Combine (A.4) and (A.5) to get

$$\|\mathbf{T}^n - \mathbf{Q}\|_{wh} = \sup_{\varphi\in\mathscr{H}, \|\varphi\|_{wh}\le 1} \|\mathbf{T}^n\varphi - \mathbf{Q}\varphi\|_{wh} \le \sup_{\varphi\in\mathscr{H}, \|\varphi\|_{wh}\le 1} \|\varphi\|_{wh}\gamma_*\rho_*^n \le \gamma_*\rho_*^n.$$

Then we have (4.11) and this completes the proof. □

PROOF OF THEOREM 2. By using Lemma 2, standard arguments involving smoothing inequalities and Fourier inversion (cf. Chapter 4 of [5]) reduce the proof to that of showing for every $\delta > 0$, $a > 0$ and $b > 1$,

$$\sup_{\delta\le|\alpha|\le n^a} \bigl|E_\pi\bigl(e^{i\alpha' S_n}\bigr)\bigr| = o(n^{-b}). \tag{A.6}$$

To prove (A.6), we follow the same idea as (3.43) of [31], letting $\zeta_t = S_t - S_{t-1}$ $(t = 1, 2, \ldots)$, $\zeta_0 = S_0$ and $\tilde{\varphi}((y,u),(y',v)) = E\{e^{i\alpha'\zeta_1}|(Y_0 = y, M_0 = u), (Y_1 = y', M_1 = v)\}$.

Let $J = \{1, \ldots, n\}$, and fix $m > 1$ to be determined later. Divide J into blocks $A_1, B_1, \ldots, A_l, B_l$ as follows. Define $j_1, \ldots, j_l$ by $j_1 = 1$, and $j_{k+1} = \inf\{j \ge j_k + 7m : j \in J\}$, and let l be the smallest integer for which the inf is undefined. Write

$$
\begin{aligned}
A_k &= \prod\bigl\{e^{n^{-1/2}i\alpha'\zeta_j} : |j - j_k| \le m\bigr\}, && k = 1, \ldots, l, \\
B_k &= \prod\bigl\{e^{n^{-1/2}i\alpha'\zeta_j} : j_k + m + 1 \le j \le j_{k+1} - m - 1\bigr\}, && k = 1, \ldots, l-1, \\
B_l &= \prod\bigl\{e^{n^{-1/2}i\alpha'\zeta_j} : j > j_l + m + 1\bigr\}.
\end{aligned}
$$

Then $e^{i\alpha' S_n} = \prod_{k=1}^{l} A_k B_k$. Given $y \in \mathcal{Y}$, we have

$$\begin{aligned}
&\left| E_y \prod_1^l A_k B_k - E_y \prod_1^l B_k E(A_k|\zeta_j : j \neq j_k) \right| \\
&\quad \le \sum_{q=1}^{l} \left| E_y \prod_1^{q-1} A_k B_k \big(A_q - E(A_k|\zeta_j : j \neq j_q)\big) \prod_{q+1}^{l} B_k E(A_k|\zeta_j : j \neq j_k) \right|.
\end{aligned} \tag{A.7}$$

By using Lemma 2(iv), there exists $\delta > 0$ such that $E|E(A_k|\zeta_j : j \neq j_q) - E(A_k|\zeta_j : 0 < |j - j_k| \le 3m)| \le e^{-\delta m}$. Therefore, (A.7) $\le$

$$\begin{aligned}
&\sum_{q=1}^{l} \left| E_y \prod_1^{q-1} A_k B_k \big(A_q - E(A_k|\zeta_j : j \neq j_q)\big) \right. \\
&\qquad \left. \times \prod_{q+1}^{l} B_k E(A_k|\zeta_j : 0 < |j - j_k| \le 3m) \right| \\
&\qquad + \sum_{q=1}^{l} e^{-\delta m}.
\end{aligned} \tag{A.8}$$

The first summation term in (A.8) vanishes since $\prod_1^{q-1} A_k B_k$ and $\prod_{q+1}^{l} B_k \times E(A_k|\zeta_j : 0 < |j - j_k| \le 3m)$ are both measurable with respect to the σ-field generated by $\zeta_j : j \neq j_q$.

Recall that the functions $E(A_k|\zeta_j : 0 < |j - j_k| \le 3m)$, for $k = 1, \dots, l$, are weakly dependent since $j_{k+1} - j_k \ge 7m$, $k = 1, \dots, l-1$. Using Assumption K1, (4.14) and (4.15), we obtain

$$\begin{aligned}
&\left| E_y \prod_1^l B_k E(A_k|\zeta_j : 0 < |j - j_k| \le 3m) \right| \\
&\quad \le E_y \left| \prod_1^l E(A_k|\zeta_j : 0 < |j - j_k| \le 3m) \right| \\
&\quad \le \prod_1^l E_y \big| E(A_k|\zeta_j : 0 < |j - j_k| \le 3m) \big| + l e^{-\delta m}.
\end{aligned}$$

With the strong nonlattice condition (4.16), and conditional strong nonlattice condition (4.17), we find an upper bound for $E_y|E(A_k|\zeta_j : 0 < |j - j_k| \le 3m)|$.

We have for $|\alpha| \ge \delta$ the relation $E_y|E(A_k|\zeta_j : j \neq j_q)| \le e^{-\delta}$ and, hence, by (4.17) for all $\alpha \in \mathbf{R}^p$, $|\alpha| \le \delta$, $E_y|E(A_k|\zeta_j : j \neq j_q)| \le \exp(-\delta|\alpha|^2/n)$. Therefore, for all $\alpha \in \mathbf{R}^p$,

$$\begin{aligned}
&E_y \big| E(A_k|\zeta_j : 0 < |j - j_k| \le 3m) \big| \\
&\quad \le e^{-\delta m} + E_y \big| E(A_k|\zeta_j : j \neq j_q) \big| \le e^{-\delta m} + \max\big(\exp(-\delta|\alpha|^2/n), e^{-\delta}\big).
\end{aligned}$$

If we choose K appropriately and let m be the integral part of $K \log n$, then the assertion of the lemma follows from $\exp(-\delta|\alpha|^2/n)^{n/m} \le \exp(-\delta|\alpha|^2/(K \log n)) \le \exp(-\delta' n^{\varepsilon/2})$ for $|\alpha| \ge cn^{\varepsilon}$ and some $\delta' > 0$. $\square$

Acknowledgments. The author is grateful to the Editor Professor Jianqing Fan, an Associate Editor and a referee for constructive comments, suggestions and correction of some errors in the earlier version.

REFERENCES

[1] BALL, F. G. and RICE, J. A. (1992). Stochastic models for ion channels: Introduction and bibliography. *Math. Biosci.* **112** 189–206.

[2] BARNSLEY, M. F., ELTON, J. H. and HARDIN, D. P. (1989). Recurrent iterated functions systems. Fractal approximation. *Constr. Approx.* **5** 3–31. MR0982722

[3] BAUM, L. E. and PETRIE, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **37** 1554–1563. MR0202264

[4] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. MR0471142

[5] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. MR0436272

[6] BICKEL, P. and RITOV, Y. (1996). Inference in hidden Markov models. I. Local asymptotic normality in the stationary case. *Bernoulli* **2** 199–228. MR1416863

[7] BICKEL, P., RITOV, Y. and RYDÉN, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* **26** 1614–1635. MR1647705

[8] BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31** 307–327. MR0853051

[9] BOLLERSLEV, T., ENGLE, R. F. and NELSON, D. B. (1994). ARCH models in finance. In *Handbook of Econometrics* **4** (R. F. Engle and D. L. McFadden, eds.) 2959–3038. North-Holland, Amsterdam. MR1315984

[10] BOUGEROL, P. and PICARD, N. (1992). Strict stationary of generalized autoregressive processes. *Ann. Probab.* **20** 1714–1730. MR1188039

[11] CAINES, P. E. (1988). *Linear Stochastic Systems*. Wiley, New York. MR0944080

[12] CLARK, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41** 135–155. MR0415944

[13] DIACONIS, P. and FREEDMAN, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737

[14] DOUC, R. and MATIAS, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7** 381–420. MR1836737

[15] DOUC, R., MOULINES, É. and RYDÉN, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.* **32** 2254–2304. MR2102510

[16] DUNSMUIR, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.* **7** 490–506. MR0527485

[17] ELLIOTT, R., AGGOUN, L. and MOORE, J. (1995). *Hidden Markov Models*: *Estimation and Control.* Springer, New York. MR1323178

[18] ELTON, J. H. (1990). A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Process. Appl.* **34** 39–47. MR1039561

[19] ENGLE, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121

[20] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods.* Springer, New York. MR1964455

[21] FRANCQ, C. and ROUSSIGNOL, M. (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum likelihood estimator. *Statistics* **32** 151–173. MR1708120

[22] FUH, C.-D. (2003). SPRT and CUSUM in hidden Markov models. *Ann. Statist.* **31** 942–977. MR1994736

[23] FUH, C.-D. (2004). On Bahadur efficiency of the maximum likelihood estimator in hidden Markov models. *Statist. Sinica* **14** 127–154. MR2036765

[24] FUH, C.-D. (2004). Uniform Markov renewal theory and ruin probabilities in Markov random walks. *Ann. Appl. Probab.* **14** 1202–1241. MR2071421

[25] FUH, C.-D. (2004). Asymptotic operating characteristics of an optimal change point detection in hidden Markov models. *Ann. Statist.* **32** 2305–2339. MR2102511

[26] FUH, C.-D. and LAI, T. L. (2001). Asymptotic expansions in multidimensional Markov renewal theory and first passage times for Markov random walks. *Adv. in Appl. Probab.* **33** 652–673. MR1860094

[27] GENON-CATALOT, V., JEANTHEAU, T. and LARÉDO, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** 1051–1079. MR1809735

[28] GHOSH, J. K. (1994). *Higher Order Asymptotics*. IMS, Hayward, CA.

[29] GHYSELS, E., HARVEY, A. C. and RENAULT, E. (1996). Stochastic volatility. In *Statistical Methods in Finance* (G. S. Maddala and C. R. Rao, eds.) 119–191. North-Holland, Amsterdam. MR1602124

[30] GOLDFELD, S. M. and QUANDT, R. E. (1973). A Markov model for switching regressions. *J. Econometrics* **1** 3–15.

[31] GÖTZE, F. and HIPP, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–239. MR0714144

[32] HALL, P. and YAO, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71** 285–317. MR1956860

[33] HAMILTON, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57** 357–384. MR0996941

[34] HAMILTON, J. D. (1994). *Time Series Analysis.* Princeton Univ. Press. MR1278033

[35] HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probab.* **10** 130–145. MR0365960

[36] HARVEY, A. C., RUIZ, E. and SHEPHARD, N. (1994). Multivariate stochastic variance models. *Rev. Econom. Stud.* **61** 247–264.

[37] HENNION, H. and HERVÉ, L. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness. Lecture Notes in Math.* **1766**. Springer, Berlin. MR1862393

[38] ITÔ, H., AMARI, S.-I. and KOBAYASHI, K. (1992). Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Trans. Inform. Theory* **38** 324–333. MR1162206

[39] JENSEN, J. L. and PETERSEN, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27** 514–535. MR1714719

[40] KROGH, A., BROWN, M., MIAN, I. S., SJÖLANDER, K. and HAUSSLER, D. (1994). Hidden Markov models in computational biology: Applications to protein modeling. *J. Molecular Biology* **235** 1501–1531.

[41] KÜNSCH, H. R. (2001). State space and hidden Markov models. In *Complex Stochastic Systems* (O. E. Barndorff-Nielsen, D. R. Cox and C. Klüppelberg, eds.) 109–173. Chapman and Hall/CRC, London. MR1893412

[42] LAHIRI, S. N. (1993). Refinements in asymptotic expansions for sums of weakly dependent random vectors. *Ann. Probab.* **21** 791–799. MR1217565

[43] LEE, S.-W. and HANSEN, B. E. (1994). Asymptotic theory for the GARCH$(1, 1)$ quasi-maximum likelihood estimator. *Econometric Theory* **10** 29–52. MR1279689

[44] LEROUX, B. G. (1992). Maximum likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* **40** 127–143. MR1145463

[45] LUMSDAINE, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH$(1, 1)$ and covariance stationary GARCH$(1, 1)$ models. *Econometrica* **64** 575–596. MR1385558

[46] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, London. MR1287609

[47] NAGAEV, S. V. (1961). More exact statement of limit theorems for homogeneous Markov chains. *Theory Probab. Appl.* **6** 62–81. MR0131291

[48] NIEMI, S. and NUMMELIN, E. (1986). On non-singular renewal kernels with an application to a semigroup of transition kernels. *Stochastic Process. Appl.* **22** 177–202. MR0860932

[49] RABINER, L. R. and JUANG, B.-H. (1993). *Fundamentals of Speech Recognition.* Prentice Hall, Englewood Cliffs, NJ.

[50] SHEPHARD, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics*, *Finance and Other Fields* (D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen, eds.) 1–67. Chapman and Hall, London.

[51] TAYLOR, S. J. (1986). *Modelling Financial Time Series*. Wiley, Chichester.

[52] TAYLOR, S. J. (1994). Modeling stochastic volatility: A review and comprehensive study. *Math. Finance* **4** 183–204.

[53] YAO, Q. and BROCKWELL, P. J. (2001). Gaussian maximum likelihood estimation for ARMA models. I. Time series. Unpublished manuscript.

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI 11529
TAIWAN
REPUBLIC OF CHINA
E-MAIL: stcheng@stat.sinica.edu.tw