



# Moderate deviation principle for autoregressive processes

Yu Miao<sup>a,\*</sup>, Si Shen<sup>b</sup>

<sup>a</sup> College of Mathematics and Information Science, Henan Normal University, 453007 Henan, China

<sup>b</sup> Department of Statistics, College of Science, Central University for Nationalities, 100081 Beijing, China

## ARTICLE INFO

### Article history:

Received 20 April 2007

Available online 21 June 2009

### 2000 MR Subject Classifications:

60F10

60G10

62J05

### Keywords:

Moderate deviation

Autoregressive processes

Least squares estimator

Yule–Walker estimator

## ABSTRACT

A moderate deviation principle for autoregressive processes is established. As statistical applications we provide the moderate deviation estimates of the least square and the Yule–Walker estimators of the parameter of an autoregressive process. The main assumption on the autoregressive process is the Gaussian integrability condition for the noise, which is weaker than the assumption of Logarithmic Sobolev Inequality in [H. Djellout, A. Guillin, L. Wu, Moderate deviations of empirical periodogram and nonlinear functionals of moving average processes, Ann. I. H. Poincaré-PR 42 (2006) 393–416].

© 2009 Elsevier Inc. All rights reserved.

## 1. Introduction

Consider the linear autoregressive model in  $\mathbb{R}^d$ ,

$$X_n = \theta X_{n-1} + \xi_n, \quad (1.1)$$

where  $\theta \in \mathcal{M}_d$  (the space of  $d \times d$  matrices) is unknown,  $(\xi_n)_{n \in \mathbb{Z}}$  is a sequence of centered i.i.d. r.v. valued in  $\mathbb{R}^d$  representing the noise and which is independent of  $X_0$ , and  $(X_n)_{n \geq 0}$  is observed. Assume that the law of  $X_0$  is invariant (or equivalently  $(X_n)_{n \geq 0}$  is stationary), there are two important issues: (i) the estimate of the covariance matrix  $\text{Cov}(X_0, X_l) := \mathbb{E}(X_0 - \mathbb{E}X_0)(X_l - \mathbb{E}X_l)^t$  (here  $X_n$  is regarded as column vector and  $A^t$  denotes the transposition of matrix  $A$ ); (ii) estimate of  $\theta$ .

It is quite easy (and well known) to check a stationary solution to (1.1), which is given by

$$X_n = \sum_{p=0}^{\infty} \theta^p \xi_{n-p}, \quad n \geq 0$$

once if  $\|\theta\| := \sup_{|x| \leq 1} |\theta x| < 1$ . So it is a special moving average process. A general moving average process is given by

$$X_n := \sum_{j=-\infty}^{+\infty} a_{j-n} \xi_j = \sum_{j=-\infty}^{+\infty} a_j \xi_{n+j}, \quad \forall n \in \mathbb{Z},$$

where  $(\xi_n)_{n \in \mathbb{Z}}$  is i.i.d.,  $(a_n)_{n \in \mathbb{Z}}$  be a sequence of real numbers such that

$$\sum_{n \in \mathbb{Z}} |a_n|^2 < \infty. \quad (1.2)$$

\* Corresponding author.

E-mail addresses: [yumiao728@yahoo.com.cn](mailto:yumiao728@yahoo.com.cn) (M. Yu), [henlen80@yahoo.com.cn](mailto:henlen80@yahoo.com.cn) (S. Si).

The most natural estimator of  $\Gamma(X_l, X_0) := (\text{Cov}(X_0^i, X_l^j))_{1 \leq i, j \leq d} (l \geq 0)$  is given by the empirical covariance (with the given sample  $(X_k)_{0 \leq k \leq n+l}$ )

$$C_{n,l}^* = \frac{1}{n} \sum_{k=1}^n X_{k+l} X_k^t, \quad (1.3)$$

and for estimating  $\theta$ , the following two estimators are widely used:

(i) Least Square Estimator:

$$\hat{\theta}_n = \left( \sum_{k=1}^n X_k X_{k-1}^t \right) \left( \sum_{k=1}^n X_{k-1} X_{k-1}^t \right)^{-1}. \quad (1.4)$$

(ii) Yule–Walker Estimator:

$$\tilde{\theta}_n = \left( \sum_{k=1}^n X_k X_{k-1}^t \right) \left( \sum_{k=0}^n X_k X_k^t \right)^{-1}. \quad (1.5)$$

In this paper we are interested in the moderate deviation behavior of the empirical covariance and of  $\hat{\theta}_n, \tilde{\theta}_n$ . There is a rich literature on the central limit theorem and iterated logarithmic law about those three estimators.

The study on large deviations and moderate deviations are relatively recent.

**Gaussian case** (i.e., the noise  $\xi$  is assumed Gaussian). This subject is opened by Donsker and Varadhan [1] who proved the level-3 large deviation principle for general stationary Gaussian processes under the continuity of the spectral function. Bryc and Dembo [2] (1993) proved for the first the large and moderate deviation principles for the empirical variance  $C_{n,0}^*$  even for general stationary Gaussian processes. Bercu, Gamboa and Rouault [3] proved the large deviation principle for  $C_{n,l}^*$ ,  $l \geq 0$  (which is much more delicate than  $C_{n,0}^*$ ) and for  $\hat{\theta}_n, \tilde{\theta}_n$ .

**Non-Gaussian case.** Wu [4] first extended Donsker–Varadhan’s theorem on large deviations of level-3 from stationary Gaussian processes to general moving average processes under the Gaussian integrability condition on the driven variable  $\xi$ . Djellout–Guillin–Wu [5] established, in the one-dimensional case (i.e.,  $d = 1$ ), moderate deviation principle for nonlinear functionals of a general moving average processes covering the case of  $C_{n,l}^*$  and for the periodogram, but under the assumption that the law of the driven random variable  $\xi$  satisfies the log-Sobolev inequality, stronger than the Gaussian integrability in [4]. The main contribution of this paper is to remove the assumption of log-Sobolev inequality on the driven variable, for the particular but important autoregression model.

For the Hilbertian autoregressive model with driven r.v.  $\xi$  satisfying the Gaussian integrability condition, in which  $\{\xi_k, X_k\}_{k \in \mathbb{Z}}$  take values in some separable Hilbert space  $H$ , Mas and Menneteau [6] established large and moderate deviation for the empirical mean  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ , and moderate deviation for the empirical variance matrix  $\frac{1}{n} \sum_{k=1}^n X_k \otimes X_k$ , where  $x \otimes y (x, y \in H)$  denotes the linear operator from  $H$  to  $H$ ,

$$x \otimes y : h \in H \rightarrow \langle x, h \rangle y,$$

extending the result of Bryc–Dembo [2] from  $\mathbb{R}^d$  to  $H$ , and especially from Gaussian case to general sub-Gaussian case. Furthermore, Menneteau [7] obtained some laws of the iterated logarithm in Hilbertian autoregressive models for the empirical covariance  $\frac{1}{n} \sum_{k=1}^n X_k \otimes X_k$ . Our main purpose is to extend this result to the MDP of the empirical covariance matrices  $(\frac{1}{n} \sum_{k=1}^n X_k \otimes X_{k+l})_{0 \leq l \leq M}$ . The difficulty in this generalization is similar for the passage from empirical variance to empirical covariance (i.e. from Bryc–Dembo [2] to Bercu, Gamboa and Rouault [3]) in the Gaussian case.

## 2. Main results

### 2.1. Assumptions

Let  $\{\xi_n\}_{n \in \mathbb{Z}}$  be a sequence of  $\mathbb{R}^d$ -valued centered i.i.d. random variables, and suppose the following conditions hold:

(C<sub>1</sub>) the moderate deviation scale  $(b_n)$  is a sequence of positive numbers satisfying  $1 \ll b_n \ll \sqrt{n}$ , i.e., as  $n \rightarrow \infty$ ,

$$b_n \rightarrow \infty; \quad \frac{b_n}{\sqrt{n}} \rightarrow 0.$$

(C<sub>2</sub>)  $\mathbb{E}\xi_0 = 0$  and  $\xi_0$  satisfies the Gaussian integrability condition, i.e., there exists  $\alpha > 0$  such that

$$\mathbb{E}e^{\alpha|\xi_0|^2} < \infty.$$

(C<sub>3</sub>)  $K_\theta := \sum_{k=0}^{\infty} \|\theta^k\| < +\infty$  where  $\|\theta\| := \sup_{|x| \leq 1} |\theta x|$ , that is, the maximal modulus of eigenvalues of  $\theta$  is  $< 1$ .

## 2.2. Main results

The space of  $d \times d$  real matrices  $\mathcal{M}_d$  is provided with the inner product  $\langle A, B \rangle := \text{tr}(AB^t)$  where  $\text{tr}(\cdot)$  is the trace. Let

$$U_{k,l} = \theta X_{k+l-1} \xi_k^t + \xi_{k+l} X_{k-1}^t + \xi_{k+l} \xi_k^t - \theta^l \Gamma(\xi_0, \xi_0), \quad \vec{U}_k := (U_{k,l})_{0 \leq l \leq M}.$$

**Theorem 2.1.** Assume that the conditions  $(C_1)$ ,  $(C_2)$  and  $(C_3)$  are satisfied, then  $\left( \frac{1}{b_n \sqrt{n}} \sum_{k=1}^n (X_{k+l} X_k^t - \mathbb{E} X_{k+l} X_k^t) \right)_{0 \leq l \leq M}$  satisfies the large deviation principle on  $\mathcal{M}_d^{M+1}$  with speed  $b_n^2$  and with the rate function given by

$$I(\Gamma) = \sup_{\Lambda \in \mathcal{M}_d^{M+1}} \left\{ \sum_{l=0}^M \langle \Lambda_l, \Gamma_l - \theta \Gamma_l \theta^t \rangle - \frac{1}{2} \Sigma^2(\Lambda) \right\}, \quad \forall \Gamma = (\Gamma_0, \dots, \Gamma_M) \in \mathcal{M}_d^{M+1}. \quad (2.1)$$

Here  $\langle \Lambda, \Gamma \rangle := \sum_{k=0}^M \langle \Lambda_k, \Gamma_k \rangle$  and

$$\Sigma^2(\Lambda) = \mathbb{E} \left( \langle \Lambda, \vec{U}_1 \rangle \right)^2 + 2 \mathbb{E} \sum_{k=2}^{M+1} \langle \Lambda, \vec{U}_1 \rangle \langle \Lambda, \vec{U}_k \rangle. \quad (2.2)$$

In particular for every  $l \geq 0$  fixed,  $\frac{1}{b_n \sqrt{n}} \sum_{k=1}^n (X_{k+l} X_k^t - \mathbb{E} X_{k+l} X_k^t)$  satisfies the large deviation principle on  $\mathcal{M}_d$  with speed  $b_n^2$  and with the rate function given by

$$I(\Gamma_l) = \sup_{\Lambda_l \in \mathcal{M}_d} \left\{ \langle \Lambda_l, \Gamma_l - \theta \Gamma_l \theta^t \rangle - \frac{1}{2} \mathbb{E} \left( \langle \Lambda_l, U_{1,l} \rangle \right)^2 \right\}, \quad \forall \Gamma_l \in \mathcal{M}_d. \quad (2.3)$$

## 2.3. Applications

In the subsection, we provide a statistical application. More precisely we shall apply [Theorem 2.1](#) to the least square estimator  $\hat{\theta}_n$  and the Yule–Walker estimator  $\tilde{\theta}_n$ :

**Proposition 2.2.** Suppose the conditions  $(C_1)$ ,  $(C_2)$  and  $(C_3)$  and assume that the covariance matrix  $\mathbb{E} X_0 X_0^t$  is non-singular. Then  $\frac{\sqrt{n}}{b_n} (\hat{\theta}_n - \theta)$  as well as  $\frac{\sqrt{n}}{b_n} (\tilde{\theta}_n - \theta)$  satisfies the large deviation principle on  $\mathcal{M}_d$  with speed  $b_n^2$  and the rate function given by

$$I(\Gamma) = \sup_{\Lambda \in \mathcal{M}_d} \left( \langle \Lambda, \Gamma \rangle - \frac{1}{2} \mathbb{E} \langle \Lambda, \xi_1 X_0^t [\mathbb{E} X_0 X_0^t]^{-1} \rangle^2 \right).$$

In particular, when  $d = 1$ , the rate function above can be identified as

$$I(x) = \frac{x^2}{2(1 - \theta^2)}.$$

**Remark 2.3.** Under our conditions it is quite easy to see that the Yule–Walker estimator  $\tilde{\theta}_n$  shares the same MDP as the least square estimator  $\hat{\theta}$ . A curious phenomena was found by Bercu, Gamboa and Rouault [3]: in the Gaussian noise and one-dimensional case, they proved the large deviations of  $\hat{\theta}$  and  $\tilde{\theta}_n$ , which possess two different rate functions. The MDP above was proved by Worms [8], where the rate function is identified basing on the Kronecker product of matrices. Djellout–Guillin–Wu [5] derived it as a consequence of their general results on the MDP of moving average processes, but with an extra and strong condition that the law of  $\xi_0$  satisfies a log-Sobolev inequality (though their method go far beyond the regression model).

## 3. Proofs

### 3.1. Proof of Theorem 2.1

Let  $m$  be a given positive integer, a sequence  $(Z_n)_{n \geq 1}$  of strictly stationary random variables is called  $m$ -dependent if for every  $k \geq 1$  the two collections  $\{Z_1, \dots, Z_k\}$  and  $\{Z_{k+m}, Z_{k+m+1}, \dots\}$  are independent. We have the following:

**Lemma 3.1** (Chen [9]). Let  $(Z_n)_{n \geq 1}$  be a stationary sequence of  $m$ -dependent random variables taking values in  $\mathbb{R}^d$ , such that

$$\mathbb{E}(e^{\alpha |Z_1|}) < \infty, \quad \text{for some } \alpha > 0.$$

Then for all  $\lambda \in \mathbb{R}^d$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \left( e^{b_n^2 \langle \lambda, \frac{1}{\sqrt{nb_n}} \sum_{k=1}^n (Z_k - \mathbb{E}Z_k) \rangle} \right) &= \frac{1}{2} \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left\langle \lambda, \sum_{k=1}^n (Z_k - \mathbb{E}Z_k) \right\rangle^2 \\ &= \frac{1}{2} \left( \mathbb{E} \langle \lambda, Z_1 \rangle^2 + 2 \sum_{k=2}^{m+1} \mathbb{E} \langle \lambda, Z_1 \rangle \langle \lambda, Z_k \rangle \right). \end{aligned}$$

*Step 1: (Autoregressive Representation for the Covariance Process)*

By the stationarity of  $\{X_n\}$ , for every  $k \in \mathbb{Z}$ , the distribution law of  $X_{k+l}X_k^t$  is the same with  $X_lX_0^t$ . Let  $C_l := \mathbb{E}X_{k+l}X_k^t = \Gamma(X_l, X_0)$  and it is easy to see that

$$C_l = \theta^l \Gamma(X_0, X_0) = \theta^l \sum_{k=0}^{\infty} \theta^k \Gamma(\xi_0, \xi_0) (\theta^k)^t = \mathbb{E}C_{n,l}^*, \quad (3.1)$$

where  $C_{n,l}^*$  is defined in (1.3). In addition, let

$$Z_{k,l} = X_{k+l}X_k^t - C_l, \quad U_{k,l} = \theta X_{k+l-1}\xi_k^t + \xi_{k+l}X_{k-1}^t\theta^t + \xi_{k+l}\xi_k^t - \theta^l \Gamma(\xi_0, \xi_0), \quad (3.2)$$

and

$$\vec{U}_k = (U_{k,l})_{0 \leq l \leq M}.$$

We have the following autoregressive representation for the covariance process.

**Lemma 3.2.** Under the above notions, for any  $k \in \mathbb{Z}$ ,  $l \in \mathbb{Z}^+ \cup \{0\}$ , we have

$$Z_{k,l} = \theta Z_{k-1,l}\theta^t + U_{k,l}, \quad (3.3)$$

and

$$C_{n,l}^* - C_l = (1 - R)^{-1}(\bar{U}_{n,l}) + (1 - R)^{-1}R \left( \frac{Z_{0,l} - Z_{n,l}}{n} \right), \quad (3.4)$$

where  $\bar{U}_{n,l} = n^{-1} \sum_{k=1}^n U_{k,l}$  and the linear operator  $R : s \in \mathcal{M}(d \times d) \mapsto \theta s \theta^t$ .

**Proof.** Easy calculus, so omitted.  $\square$

The following result shows that the main part of  $C_{n,l}^* - C_l$  is  $(1 - R)^{-1}\bar{U}_{n,l}$  in the sense of moderate deviation.

**Lemma 3.3.** Under the conditions  $(C_1)$  and  $(C_2)$ , there exists  $\alpha > 0$  such that, for all  $r > 0$ ,

$$\mathbb{P}(\|Z_{0,l} - Z_{n,l}\|_{HS} > rb_n\sqrt{n}) \leq 4 \exp\left(-\alpha \frac{rb_n\sqrt{n}}{2K_\theta^2}\right) \mathbb{E}(e^{\alpha|\xi_0|^2}), \quad (3.5)$$

where  $K_\theta$  is given by

$$K_\theta := \sum_{k=0}^{\infty} \|\theta^k\|. \quad (3.6)$$

Consequently for any  $r > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n} \|(1 - R)^{-1}R(Z_{0,l} - Z_{n,l})\|_{HS}}{n} > r\right) = -\infty. \quad (3.7)$$

Here  $\|A\|_{HS} = \sqrt{\langle A, A \rangle} = \sqrt{\text{tr}(AA^t)}$  is the Hilbert–Schmidt norm.

**Proof.** From the stationarity of  $\{X_n\}_{n \geq 0}$ , we get

$$\begin{aligned} \mathbb{P}(\|Z_{0,l} - Z_{n,l}\|_{HS} > rb_n\sqrt{n}) &= \mathbb{P}(\|X_lX_0^t - X_{n+l}X_n^t\|_{HS} > rb_n\sqrt{n}) \\ &\leq 2\mathbb{P}(\|X_lX_0^t\|_{HS} > rb_n\sqrt{n}/2) \\ &\leq 4\mathbb{P}(|X_0|^2 > rb_n\sqrt{n}/2) \end{aligned}$$

where the last inequality follows from Cauchy–Schwarz:  $\|X_lX_0^t\|_{HS} \leq \frac{1}{2}(|X_l|^2 + |X_0|^2)$ . Since the remainders of the proof is similar to Lemma 13 of Mas and Menneteau in [6], here we omit them.  $\square$

**Step 2: (Asymptotic Term and Moderate Deviation)**

For all  $k \geq 1$  and  $m \geq 2$ ,  $m > l$ , set

$$\begin{aligned} X_{k-1,m} &= \xi_{k-1} + \theta \xi_{k-2} + \cdots + \theta^{m-2} \xi_{k-m+1} = \sum_{j=0}^{m-2} \theta^j \xi_{k-1-j}, \\ U_{k,l,m} &= \theta X_{k+l-1,m} \xi_k^t + \xi_{k+l} X_{k-1,m}^t \theta^t + \xi_{k+l} \xi_k^t - \theta^l \Gamma(\xi_0, \xi_0) \\ &= \sum_{j=1}^{m-1} \theta^j \xi_{k+l-j} \xi_k^t + \sum_{j=1}^{m-1} \xi_{k+l} \xi_{k-j}^t (\theta^j)^t + \xi_{k+l} \xi_k^t - \theta^l \Gamma(\xi_0, \xi_0), \end{aligned}$$

and

$$\vec{U}_{k,m} = (U_{k,l,m})_{0 \leq l \leq M}.$$

It is easy to see that  $\{\vec{U}_{k,m}\}_{k \geq 1}$  is a strictly stationary sequence with  $(M+m)$ -dependent structure. Furthermore for each  $l \geq 0$  fixed,  $\{U_{k,l,m}, k \geq 0\}$  is a martingale difference sequence w.r.t.  $(\mathcal{F}_k)_{k \geq 0}$ , where  $\mathcal{F}_n := \sigma(\xi_k; -\infty < k \leq n)$ . Set

$$\bar{U}_{n,l,m} = \frac{1}{n} \sum_{k=1}^n U_{k,l,m} \quad \text{and} \quad Q_{n,m} = (\bar{U}_{n,l,m})_{0 \leq l \leq M}.$$

Applying Chen's Lemma 3.1, we have

**Lemma 3.4.** Under the conditions  $(C_1)$ ,  $(C_2)$  and  $(C_3)$ , for all  $m \geq 1$ ,  $\Lambda = (\Lambda_0, \dots, \Lambda_M) \in \mathcal{M}_d^{M+1}$ , writing  $\langle \Lambda, Q_{n,m} \rangle := \sum_{l=0}^M \langle \Lambda_l, \bar{U}_{n,l,m} \rangle$  we have

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \exp \{b_n \sqrt{n} \langle \Lambda, Q_{n,m} \rangle\} = \frac{1}{2} \Sigma_m^2(\Lambda), \quad (3.8)$$

where  $\Sigma_m^2(\Lambda) := \mathbb{E} \langle \Lambda, \vec{U}_{1,m} \rangle^2 + 2 \sum_{k=2}^{M+1} \mathbb{E} \langle \Lambda, \vec{U}_{1,m} \rangle \cdot \langle \Lambda, \vec{U}_{k,m} \rangle$ .

**Step 3: (The Asymptotic Negligibility of  $\frac{1}{b_n \sqrt{n}} \{\vec{U}_n - \vec{U}_{n,m}\}_{n \geq 1}$  as  $m \rightarrow \infty$ )**

This is, of course, the crucial and difficult step. Without loss of generality, we only consider  $\{\bar{U}_{n,l} - \bar{U}_{n,l,m}\}_{n \geq 1}, \forall 0 \leq l \leq M$ . In the next result, we establish an exponential inequality for  $\{\bar{U}_{n,l} - \bar{U}_{n,l,m}\}_{n \geq 1}$  and we obtain that  $\left\{ \frac{\sqrt{n}}{b_n} \bar{U}_{n,l,m} \right\}_{n \geq 1, m \geq 2}$  is a  $b_n^2$ -exponentially good approximation of the sequence  $\left\{ \frac{\sqrt{n}}{b_n} \bar{U}_{n,l} \right\}_{n \geq 1}$ .

For all  $p \geq 0$  and  $k \geq 1$ , set

$$W_{k,p} = \xi_k \left( \frac{\theta^p}{\|\theta^p\|} \xi_{k-p} \right)^t.$$

**Lemma 3.5.** Assume the conditions  $(C_1)$ ,  $(C_2)$  and  $(C_3)$ .

(i) There exist  $\alpha_0$  and  $\beta_0$  such that, for all  $p \geq 1$ ,  $n \geq 1$  and  $t \geq 0$ ,

$$\mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j W_{k,p} \right\|_{HS} \geq t \right) \leq 36 \exp \left( - \frac{t^2}{\alpha_0 n + \beta_0 t} \right). \quad (3.9)$$

(ii) For all  $r > 0$ , there exist  $N \geq 1$  and  $A, B > 0$  such that, for all  $n \geq N$  and  $m \geq 1$ ,

$$\begin{aligned} &\mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j (U_{k,l} - U_{k,l,m}) \right\|_{HS} > r b_n \sqrt{n} \right) \\ &\leq 72 \left( 1 - \exp \left( - \frac{b_n^2 r^2}{(Ar + B) \|\theta^m\|} \right) \right)^{-1} \exp \left( - \frac{r^2 b_n^2}{(Ar + B) \|\theta^m\|} \right). \end{aligned} \quad (3.10)$$

(iii) For all  $r > 0$ ,

$$\limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{\sqrt{n}}{b_n} \|\bar{U}_{n,l} - \bar{U}_{n,l,m}\|_{HS} > r \right) = -\infty.$$

(iv) For every  $\lambda > 0$ ,

$$\limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \exp (\lambda \sqrt{n} b_n \|\bar{U}_{n,l} - \bar{U}_{n,l,m}\|_{HS}) = 0.$$

**Proof.** (i) This is Lemma 17 in [6].

(ii) The approach to prove the inequality stems from Lemma 17 of Mas and Menneteau in [6]. For the sake of completeness, we state the proof. We have

$$\begin{aligned} \left\| \sum_{k=1}^n (U_{k,l} - U_{k,l,m}) \right\|_{HS} &= \left\| \sum_{k=1}^n [\theta(X_{k+l-1} - X_{k+l-1,m})\xi_k^t + \xi_{k+l}(X_{k-1} - X_{k-1,m})^t \theta^t] \right\|_{HS} \\ &\leq \left\| \sum_{k=1}^n \theta(X_{k+l-1} - X_{k+l-1,m})\xi_k^t \right\|_{HS} + \left\| \sum_{k=1}^n \xi_{k+l}(X_{k-1} - X_{k-1,m})^t \theta^t \right\|_{HS}. \end{aligned} \quad (3.11)$$

Moreover,

$$X_{k+l-1} - X_{k+l-1,m} = \sum_{p=m-1}^{\infty} \theta^p \xi_{k+l-1-p} = \theta^{m-1} \left( \sum_{p=0}^{\infty} \theta^p \xi_{k+l-m-p} \right).$$

Hence using  $\|AB\|_{HS} \leq \|A\| \|B\|_{HS}$ , we have

$$\begin{aligned} \left\| \sum_{k=1}^n \theta(X_{k+l-1} - X_{k+l-1,m})\xi_k^t \right\|_{HS} &= \left\| \sum_{p=0}^{\infty} \sum_{k=1}^n \theta^m (\theta^p \xi_{k+l-m-p}) \xi_k^t \right\|_{HS} \\ &\leq \|\theta^m\| \sum_{p=0}^{\infty} \|\theta^p\| \left\| \sum_{k=1}^n W_{k,m+p-l} \right\|_{HS}. \end{aligned}$$

Now we need the following fact: under  $(C_3)$ ,

$$K_1 := \sum_{p=0}^{\infty} (p+1) \|\theta^p\| < \infty.$$

Therefore, if we set

$$t_{m,p}(r) = \frac{r(p+1)}{2K_1 \|\theta^m\|},$$

we have, by (3.9),

$$\begin{aligned} \mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j \theta(X_{k+l-1} - X_{k+l-1,m})\xi_k^t \right\|_{HS} > rb_n \sqrt{n}/2 \right) \\ \leq \mathbb{P} \left( \sum_{p=0}^{\infty} (p+1) \frac{\|\theta^p\|}{p+1} \max_{j \leq n} \left\| \sum_{k=1}^j W_{k,m+p-l} \right\|_{HS} > \sum_{p=0}^{\infty} (p+1) \|\theta^p\| \frac{rb_n \sqrt{n}}{2K_1 \|\theta^m\|} \right) \\ \leq \sum_{p=0}^{\infty} \mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j W_{k,m+p-l} \right\|_{HS} > \frac{(p+1)rb_n \sqrt{n}}{2K_1 \|\theta^m\|} \right) \\ \leq 36 \sum_{p=0}^{\infty} \exp \left( - \frac{b_n^2 t_{m,p}^2(r)}{\alpha_0 + \beta_0 t_{m,p}(r) b_n / \sqrt{n}} \right). \end{aligned} \quad (3.12)$$

By the assumption of  $b_n$ , there exists constants  $N \in \mathbb{N}^*$ ,  $A, B > 0$ , such that for all  $n \geq N$ ,  $m \geq 1$  and  $l \geq 0$ ,  $\frac{\sqrt{n}}{b_n} \geq 1$  and we obtain

$$\frac{t_{m,p}^2(r)}{\alpha_0 + \beta_0 t_{m,p}(r) b_n / \sqrt{n}} \geq c(r) \frac{p+1}{\|\theta^m\|}, \quad c(r) := \frac{r^2}{Ar + B}. \quad (3.13)$$

Hence, by (3.12) and (3.13), we get

$$\mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j \theta(X_{k+l-1} - X_{k+l-1,m})\xi_k^t \right\|_{HS} > rb_n \sqrt{n}/2 \right) \leq 36 \sum_{p=0}^{\infty} \exp \left( -b_n^2 \frac{c(r)}{\|\theta^m\|} (p+1) \right). \quad (3.14)$$

For the same reason, we have

$$\begin{aligned} \left\| \sum_{k=1}^n \xi_{k+l}(X_{k-1} - X_{k-1,m})^t \theta^t \right\|_{HS} &= \left\| \sum_{p=0}^{\infty} \sum_{k=1}^n \xi_{k+l} \xi_{k-m-p}^t (\theta^{p+m})^t \right\|_{HS} \\ &\leq \|\theta^m\| \sum_{p=0}^{\infty} \|\theta^p\| \left\| \sum_{k=1}^n W_{k+l,m+p+l} \right\|_{HS}, \end{aligned}$$

and for all  $n \geq N$ ,

$$\mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j \xi_{k+l}(X_{k-1} - X_{k-1,m})^t \theta^t \right\|_{HS} > rb_n \sqrt{n}/2 \right) \leq 36 \sum_{p=0}^{\infty} \exp \left( -b_n^2 \frac{c(r)}{\|\theta^m\|} (p+1) \right). \quad (3.15)$$

So, from (3.14) and (3.15), we obtain

$$\begin{aligned} \mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j (U_{k,l} - U_{k,l,m}) \right\|_{HS} > rb_n \sqrt{n} \right) &\leq \mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j \theta(X_{k+l-1} - X_{k+l-1,m}) \xi_k^t \right\|_{HS} > rb_n \sqrt{n}/2 \right) \\ &\quad + \mathbb{P} \left( \max_{j \leq n} \left\| \sum_{k=1}^j \xi_{k+l}(X_{k-1} - X_{k-1,m})^t \theta^t \right\|_{HS} > rb_n \sqrt{n}/2 \right) \\ &\leq 72 \sum_{p=0}^{\infty} \exp \left( -b_n^2 \frac{c(r)}{\|\theta^m\|} (p+1) \right) = K(r) \exp \left( -b_n^2 \frac{c(r)}{\|\theta^m\|} \right) \end{aligned}$$

where

$$K(r) = 72 \left( 1 - \exp \left( -\frac{b_n^2 c(r)}{\|\theta^m\|} \right) \right)^{-1}$$

the desired inequality.

(iii) It follows obviously by (3.10).

(iv) The inequality  $\geq$  in (iv) is obvious. Below we prove the converse inequality. Let

$$Z = Z_{n,m} := \frac{1}{b_n \sqrt{n}} \left\| \sum_{k=1}^n (U_{k,l} - U_{k,l,m}) \right\|_{HS}$$

and  $\delta > 0$  be an arbitrary positive constant. We have

$$\begin{aligned} \mathbb{E} e^{\lambda b_n^2 Z_{n,m}} &= \mathbb{E} \int_0^Z \lambda b_n^2 e^{\lambda b_n^2 r} dr + 1 = 1 + \int_0^\infty \lambda b_n^2 e^{\lambda b_n^2 r} \mathbb{P}(Z > r) dr \\ &\leq 1 + \lambda b_n^2 e^{\lambda b_n^2 \delta} + \int_\delta^\infty \lambda b_n^2 e^{\lambda b_n^2 r} \mathbb{P}(Z > r) dr. \end{aligned}$$

Choosing  $m_0 \geq 1$  so that  $\frac{r}{(Ar+B)\|\theta^m\|} \geq \lambda + 1$  for all  $r \geq \delta$ , we have by (ii),

$$\begin{aligned} \mathbb{P}(Z_{n,m} > r) &\leq 72 \left( 1 - \exp \left( -\frac{b_n^2 r^2}{(Ar+B)\|\theta^m\|} \right) \right)^{-1} \exp \left( -b_n^2 \frac{r^2}{(Ar+B)\|\theta^m\|} \right) \\ &\leq C(\delta) \exp(-(\lambda+1)b_n^2 r) \end{aligned}$$

where  $C(\delta)$  is some constant (independent of  $m, n$ ). Consequently

$$\int_\delta^\infty \lambda b_n^2 e^{\lambda b_n^2 r} \mathbb{P}(Z > r) dr \leq \lambda b_n^2 \int_\delta^\infty C(\delta) e^{-b_n^2 r} dr = \lambda C(\delta) e^{-b_n^2 \delta} \leq \lambda C(\delta).$$

Thus

$$\begin{aligned} \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \exp(\lambda \sqrt{n} b_n \|\bar{U}_{n,l} - \bar{U}_{n,l,m}\|_{HS}) &\leq \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \left( 1 + \lambda b_n^2 e^{\lambda b_n^2 \delta} + \lambda C(\delta) \right) \\ &= \lambda \delta. \end{aligned}$$

As  $\delta > 0$  is arbitrary, we obtain the desired claim.  $\square$

**Step 4: (The Identification of Rate Function)**

At first  $\Sigma_m^2(\Lambda) \rightarrow \Sigma^2(\Lambda)$  as  $m$  goes to infinity, where  $\Sigma_m^2(\Lambda)$  is given in [Lemma 3.4](#), and  $\Sigma^2(\Lambda)$  is given in [Theorem 2.1](#). For every  $k \geq M + 2$ , by considering the case:  $l = 0$  and  $0 < l \leq M$  separately, we obtain

$$\mathbb{E}(U_{k,l}|\mathcal{F}_{1+M}) = \mathbb{E}[\theta X_{k+l-1}\xi_k^t + \xi_{k+l}X_{k-1}^t\theta^t + \xi_{k+l}\xi_k^t - \theta^l\Gamma(\xi_0, \xi_0)|\mathcal{F}_{1+M}] = 0.$$

From the properties of conditional expectation,

$$\mathbb{E}U_{1,i}U_{k,l} = \mathbb{E}[U_{1,i}\mathbb{E}(U_{k,l}|\mathcal{F}_{1+M})] = 0,$$

for any  $0 \leq i, l \leq M$  and  $k \geq M + 2$ .

Let  $Q_n = \{(\bar{U}_{n,l})_{0 \leq l \leq M}\}$ , we will establish the moderate deviation of  $Q_n$ , namely, for any  $\Lambda \in \mathcal{M}_d^{M+1}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \exp \{b_n \sqrt{n} \langle \Lambda, Q_n \rangle\} = \frac{1}{2} \Sigma^2(\Lambda). \quad (3.16)$$

For any fixed  $p, q > 1$  with  $\frac{1}{p} + \frac{1}{q} = 1$ , by the Hölder inequality we have that

$$\log \mathbb{E} \exp \{b_n \sqrt{n} \langle \Lambda, Q_n \rangle\} \leq \frac{1}{p} \log \mathbb{E} \exp \{pb_n \sqrt{n} \langle \Lambda, Q_{n,m} \rangle\} + \frac{1}{q} \log \mathbb{E} \exp \{qb_n \sqrt{n} \langle \Lambda, Q_n - Q_{n,m} \rangle\}$$

for all  $\Lambda \in \mathcal{M}_d^{M+1}$ . From [Lemma 3.4](#) and (iv) in [Lemma 3.5](#), we have

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{E} \exp \{b_n \sqrt{n} \langle \Lambda, Q_n \rangle\} \leq \frac{p}{2} \Sigma^2(\Lambda). \quad (3.17)$$

Similarly, by the Hölder inequality, we have for every  $\Lambda$ ,

$$\begin{aligned} & \frac{1}{b_n^2} \log \mathbb{E} \exp \{p^{-1}b_n \sqrt{n} \langle \Lambda, Q_{n,m} \rangle\} \\ & \leq \frac{1}{b_n^2} \left( \frac{1}{p} \log \mathbb{E} \exp \{b_n \sqrt{n} \langle \Lambda, Q_n \rangle\} + \frac{1}{q} \log \mathbb{E} \exp \{(q/p)b_n \sqrt{n} \langle \Lambda, Q_{n,m} - Q_n \rangle\} \right). \end{aligned}$$

Taking first  $\liminf_{n \rightarrow \infty}$  and next  $\lim_{m \rightarrow \infty}$  we get from [Lemma 3.4](#) and (iv) in [Lemma 3.5](#)

$$\frac{1}{2p^2} \Sigma^2(\Lambda) \leq \liminf_{n \rightarrow \infty} \frac{1}{b_n^2} \mathbb{E} \exp \{b_n \sqrt{n} \langle \Lambda, Q_n \rangle\}. \quad (3.18)$$

Letting  $p \rightarrow 1$  in (3.17) and (3.18) yields (3.16).

By the Ellis–Gärtner theorem ([10], Section 2.3), (3.16) implies that  $\mathbb{P}\left(\frac{\sqrt{n}}{b_n}Q_n \in \cdot\right)$  satisfies the large deviation principle on  $\mathcal{M}_d^{M+1}$  with speed  $b_n^2$  and with the rate function given by

$$J(\Gamma) = \sup_{\Lambda \in \mathcal{M}_d^{M+1}} \left( \langle \Lambda, \Gamma \rangle - \frac{1}{2} \Sigma^2(\Lambda) \right).$$

By [Lemmas 3.2](#) and [3.3](#) and the contraction principle,  $\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(C_{n,l}^* - C_l) \in \cdot\right)$  satisfies the large deviation principle on  $\mathcal{M}_d^{M+1}$  with speed  $b_n^2$  and with the rate function

$$I(\Gamma) = J((1 - R)\Gamma)$$

which is exactly the expression (2.1) of  $I$  by the definition of  $J$  above.

**3.2. Proof of Proposition 2.2**

Let us introduce

$$r_n := \frac{\sqrt{n}}{b_n}(\hat{\theta}_n - \theta) \quad \text{and} \quad R_n = \frac{1}{\sqrt{n}b_n} \sum_{i=1}^n (X_i X_{i-1}^t - \theta X_{i-1} X_{i-1}^t)(\mathbb{E}X_0 X_0^t)^{-1}.$$

By [Theorem 2.1](#),  $R_n$  satisfies the moderate deviation principle. Before identifying its rate function let us first show that  $r_n - R_n$  is negligible with respect to the moderate deviation principle, i.e., for any  $r > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\|r_n - R_n\|_{HS} > r) = -\infty. \quad (3.19)$$



To that end, note

$$\begin{aligned} r_n &= \frac{\sqrt{n}}{b_n} \left( \sum_{i=1}^n (X_i X_{i-1}^t - \theta X_{i-1} X_{i-1}^t) \right) \left( \sum_{i=1}^n X_{i-1} X_{i-1}^t \right)^{-1} \\ &= \frac{1}{b_n \sqrt{n}} \left( \sum_{i=1}^n (X_i X_{i-1}^t - \theta X_{i-1} X_{i-1}^t) \right) \times n \times \left( \sum_{i=1}^n X_{i-1} X_{i-1}^t \right)^{-1}. \end{aligned}$$

Thus

$$r_n - R_n = \frac{1}{b_n \sqrt{n}} \left( \sum_{i=1}^n (X_i X_{i-1}^t - \theta X_{i-1} X_{i-1}^t) \right) (\mathbb{E} X_0 X_0^t)^{-1} \times \left( \mathbb{E} X_0 X_0^t - \frac{1}{n} \sum_{i=1}^n X_{i-1} X_{i-1}^t \right) \times n \times \left( \sum_{i=1}^n X_{i-1} X_{i-1}^t \right)^{-1}.$$

For any  $r > 0$ ,  $L > 0$  and  $\delta > 0$ , using  $\|AB\|_{HS} \leq \|A\|_{HS} \|B\|$  we have

$$\begin{aligned} \mathbb{P}(\|r_n - R_n\|_{HS} > r) &\leq \mathbb{P}\left(\left\| \frac{1}{b_n \sqrt{n}} \sum_{i=1}^n (X_i X_{i-1}^t - \theta X_{i-1} X_{i-1}^t) (\mathbb{E} X_0 X_0^t)^{-1} \right\|_{HS} \geq L \sqrt{\delta r}\right) \\ &\quad + \mathbb{P}\left(\left\| \mathbb{E} X_0 X_0^t - \frac{1}{n} \sum_{i=1}^n X_{i-1} X_{i-1}^t \right\|_{HS} \geq \frac{\sqrt{\delta r}}{L}\right) + \mathbb{P}\left(\left\| \left( \frac{1}{n} \sum_{i=1}^n X_{i-1} X_{i-1}^t \right)^{-1} \right\| > \frac{1}{\delta}\right). \end{aligned}$$

For  $\delta, r$  sufficiently small but fixed, the first term at the r.h.s. above is negligible by the moderate deviation principle of  $R_n$  by letting  $L \rightarrow \infty$ . The second one is bounded from above by (for  $n$  large enough)

$$\mathbb{P}\left(\frac{1}{\sqrt{n} b_n} \left\| \sum_{i=1}^n (X_{i-1} X_{i-1}^t - \mathbb{E} X_0 X_0^t) \right\|_{HS} \geq \frac{\sqrt{n}}{b_n}\right)$$

which is clearly negligible by the moderate deviation principle of

$$\frac{1}{\sqrt{n} b_n} \sum_{i=1}^n (X_{i-1} X_{i-1}^t - \mathbb{E} X_0 X_0^t).$$

For the third term, as  $\mathbb{E}(X_0 X_0^t)$  is non-degenerate, then there is some  $\delta_0 > 0$  such that for any  $d \times d$  matrix  $A$  such that  $\|A - \mathbb{E}(X_0 X_0^t)\| \leq \delta_0$ ,  $A^{-1}$  exists and  $\|A^{-1}\| \leq 2\|(\mathbb{E}(X_0 X_0^t))^{-1}\|$ . Thus for  $\delta > 0$  so that  $2\|(\mathbb{E}(X_0 X_0^t))^{-1}\| < \frac{1}{\delta}$ ,

$$\mathbb{P}\left(\left\| \left( \frac{1}{n} \sum_{i=1}^n X_{i-1} X_{i-1}^t \right)^{-1} \right\| > \frac{1}{\delta}\right) \leq \mathbb{P}\left(\left\| \frac{1}{n} \sum_{i=1}^n X_{i-1} X_{i-1}^t - \mathbb{E} X_0 X_0^t \right\| > \delta_0\right)$$

where the r.h.s. is negligible by the same argument as for the second term.

In conclusion we have proven (3.19), and so  $r_n$  satisfies the same moderate deviation principle as  $R_n$ . It remains to identify the rate function governing the moderate deviation principle of  $R_n$ . Noting that  $\sum_{i=1}^n (X_i X_{i-1}^t - \theta X_{i-1} X_{i-1}^t) (\mathbb{E} X_0 X_0^t)^{-1}$  is a martingale with stationary differences, by the similar proof of Theorem 2.1,  $R_n$  satisfies the MDP with the speed  $b_n^2$  and with the rate function

$$I(\Gamma) = \sup_{\Lambda \in \mathcal{M}_d} \left( \langle \Lambda, \Gamma \rangle - \frac{1}{2} \mathbb{E} \langle \Lambda, (X_1 X_0^t - \theta X_0 X_0^t) (\mathbb{E} X_0 X_0^t)^{-1} \rangle^2 \right). \quad (3.20)$$

But  $X_1 - \theta X_0 = \xi_1$ , so the expression above coincides with the claimed rate function.

When  $d = 1$ , then it is easy to see that

$$\mathbb{E}(X_0^2) = \frac{1}{1 - \theta^2} \mathbb{E} \xi_0^2$$

and

$$\mathbb{E}(X_1 X_0 - \theta X_0^2)^2 = \mathbb{E}(\xi_1 X_0)^2 = \mathbb{E}(\xi_0^2)^2 \frac{1}{1 - \theta^2},$$

then from (3.20), we have

$$I(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - \frac{1}{2} \lambda^2 (1 - \theta^2) \right\} = \frac{x^2}{2(1 - \theta^2)}$$

which is the desired result.

## Acknowledgments

Here the authors want to thank their supervisor – Professor Liming Wu of Université Blaise Pascal and Wuhan University – for many suggestions and helpful discussions during the preparation of the paper. Furthermore, the authors are very grateful to the referee for his/her valuable reports which improved the presentation of this work.

## References

- [1] M.D. Donsker, S.R.S. Varadhan, Large deviations for stationary processes, *Comm. Math. Phys.* 97 (1985) 187–210.
- [2] W. Bryc, A. Dembo, Large deviations for quadratic functionals of Gaussian processes, *J. Theoret. Probab.* 10 (1997) 307–322.
- [3] B. Bercu, F. Gamboa, A. Rouault, Large deviations for quadratic forms of stationary Gaussian processes, *Stochastic Process. Appl.* 71 (1997) 75–90.
- [4] L.M. Wu, On large deviations for moving average processes, in: T.L. Lai, H.L. Yang, S.P. Yung. (Eds.), *Probability, Finance and Insurance*, in: *Proceeding of a Workshop at the University of Hong Kong (15–17 July 2002)*, World Scientific, Singapore, 2004, pp. 15–49.
- [5] H. Djellout, A. Guillin, L. Wu, Moderate deviations of empirical periodogram and non-linear functionals of moving average processes, *Ann. Inst. H. Poincaré-PR* 42 (2006) 393–416.
- [6] A. Mas, L. Menneveau, Large and moderate deviations for infinite-dimensional autoregressive processes, *J. Multivariate Anal.* 87 (2003) 241–260.
- [7] L. Menneveau, Some laws of the iterated logarithm in Hilbertian autoregressive models, *J. Multivariate Anal.* 92 (2005) 405–425.
- [8] J. Worms, Moderate deviations for stable Markov chains and regression models, *Electron. J. Probab.* 4 (1999) 1–28, paper 8.
- [9] X. Chen, Moderate deviations for  $m$ -dependent random variables with Banach space values, *Statist. Probab. Lett.* 35 (1997) 123–134.
- [10] A. Dembo, O. Zeitouni, *Large deviations Techniques and Applications*, second edition, Springer, New York, 1998.