Stability of block-triangular stationary random matrices

László Gerencsér\textsuperscript{a,∗}, György Michaletzky\textsuperscript{a,b}, Zsanett Orlovits\textsuperscript{a,c}

\textsuperscript{a} MTA SZTAKI, Kende u. 13-17, 1111 Budapest, Hungary
\textsuperscript{b} Department of Probability Theory and Statistics, Eötvös Loránd University (ELTE) Pázmány Péter sétány 1/C, 1117 Budapest, Hungary
\textsuperscript{c} Department of Differential Equations, Budapest University of Technology and Economics (BME), Műegyetem rkp. 3-9, 1111 Budapest, Hungary

Received 11 April 2006; received in revised form 9 November 2007; accepted 7 January 2008
Available online 5 March 2008

Abstract

The objective of this note is to prove, under certain technical conditions, that the top-Lyapunov exponent of a strictly stationary random sequence of block-triangular matrices is equal to the maximum of the top-Lyapunov exponents of its diagonal blocks. This study is partially motivated by a basic technical problem in the identification of GARCH processes. A recent extension of the above inheritance theorem in the context of $L_q$-stability will also be briefly described.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Random coefficient linear stochastic systems; Identification; Lyapunov exponents; Random matrix products; GARCH processes

1. Introduction and the main result

Consider a linear stochastic system given by the state-space equation of the form

\[ X_{n+1} = A_{n+1}X_n + u_{n+1}, \quad -\infty < n < +\infty, \]

where $X_n \in \mathbb{R}^p$, and $(A_n, u_n)$ is a jointly strictly stationary sequence of random matrices of size $p \times (p+1)$ over some probability space $(\Omega, \mathcal{F}, P)$, where $\Omega$ is the space or set of elementary events denoted by $\omega$, the $\sigma$-algebra $\mathcal{F}$ is the set of measurable subsets of $\Omega$, and $P$ is a probability measure on $\mathcal{F}$. A strictly stationary solution $(X_n)$ is called causal if $X_n$ is measurable with respect to the $\sigma$-field $\sigma\{A_i, u_i, i \leq n\}$. Both necessary and sufficient conditions for the existence of a strictly stationary causal solution of (1) have been given in [6]. The proof of necessity is far from simple. To formulate a sufficient condition we need the concept of a Lyapunov exponent. Let $A = (A_n)$ be a strictly stationary, ergodic sequence of $p \times p$ random matrices over $(\Omega, \mathcal{F}, P)$ such that

\[ \log^+ \|A_n\| < +\infty, \]

where $\log^+ x$ denotes the positive part of $\log x$. Then it easy to see that

\[ \lambda = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \log \|A_n \ldots A_1\| \]

exists, where $-\infty \leq \lambda < +\infty$. The number $\lambda$ is called the top-Lyapunov exponent of $A$, and is denoted by $\lambda(A)$. If $A_n = A$ for all $n$ then $\lambda(A)$ is simply the logarithm of the spectral radius of $A$.

A major result of the theory of random matrices is the Fürstenberg–Kesten theorem stating that

\[ \lambda(A) = \lim_{n \to \infty} \frac{1}{n} \log \|A_n \ldots A_1\| \]

almost surely, see [11]. It follows that for all $\epsilon > 0$ the random variable

\[ c_A(\omega) = \sup_{n \geq 0} (\log \|A_n \ldots A_1\| - n(\lambda + \epsilon)) \]

exists and is finite almost surely, and thus

\[ \|A_n \ldots A_1\| \leq C_A(\omega) e^{\lambda(\lambda + \epsilon)}, \]

with $C_A(\omega) = e^{c_A(\omega)}$.

In the case when the range of $(A_n)$ is bounded, a simple crude upper bound of $\lambda(A)$ is the so-called joint spectral radius of $A$. For its computation a sequence of useful results has been given in [2,4].
Now assume that \((A_n, u_n)\) is a strictly stationary, ergodic sequence of \(p \times (p + 1)\) random matrices such that (2) holds. Assume in addition that a similar condition holds for \((u_n, \theta)\), i.e.

\[ E \log^+ \| u_n \| < +\infty. \]  

(4)

Then it is not difficult to show, see [6], that if the top-Lyapunov exponent \(\lambda(A)\) is negative then (1) has a unique strictly stationary, causal solution given by

\[ X_n^* = u_n + \sum_{k=1}^{\infty} A_n A_{n-1} \ldots A_{n-k+1} u_{n-k}. \]

(5)

Furthermore, solving (1) with any initial condition \(X_0\) forward in time we get that for any \(\varepsilon > 0\) we have with probability 1 (w.p.1)

\[ X_n - X_n^* = O(e^{(\lambda+\varepsilon)n}). \]

(6)

A nice application of linear stochastic systems with random system matrices is in the analysis of the celebrated GARCH model, which has been developed to describe stochastic volatility processes. It is well known for its capability to capture the phenomenon called volatility clustering; see [9,5]. Letting \((y_n)\) denote the log-return of a stock on day \(n\), it is assumed that

\[ y_n = \sigma_n \epsilon_n, \quad n \in \mathbb{Z}, \]

(7)

where \((\epsilon_n)\) is a sequence of i.i.d. random variables with \(E \epsilon_0 = 0\) and \(E \epsilon_0^2 = 1\), and \(\sigma_n^2\) is defined via the feedback

\[ \sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i^* y_{n-i} + \sum_{j=1}^{q} \beta_j^* \sigma_{n-j}^2, \]

(8)

where \(\alpha_0^* > 0\) and \(\alpha_i^* \geq 0\), \(\beta_j^* \geq 0\).

Now (7) and (8) can be written in a state-space form with state vector \(X_n = (y_n^2, \ldots, y_{n-p+1}^2, \sigma_n^2, \ldots, \sigma_{n-q+1}^2)^\top\). It is easy to see that \((X_n)\) satisfies a linear state-space equation of the form (1) with \((A_n, u_n)\) being an i.i.d. sequence satisfying (2), such that their elements depend linearly on the system parameters. Another application of random matrices in finance is given in [14].

Now assume that the matrices \((A_n, u_n)\) depend on a parameter \(\theta \in D \subset \mathbb{R}^k\), where \(D\) is an open set, and assume that \((A_n(\theta), u_n(\theta))\) are \(C^1\)-functions of \(\theta\) for all \(\omega \in \Omega\). Assume that the conditions of [6] described above are satisfied. Then the unique strictly stationary, causal solution of (1) will be denoted by \((X_n(\theta))\). In system identification we typically need to differentiate a parameter-dependent state-vector. It is not obvious at all that \(X_n(\theta)\) is differentiable for all \(\omega\) or for almost all \(\omega \in \Omega\). Let \(X_{\theta,n}\) denote the derivative of \(X_n(\theta)\) with respect to the parameter vector \(\theta\). Carrying out formal differentiation and assuming that \(\theta\) is scalar we get for the extended state vector \((X_n, X_{\theta,n})\) the state-space equation

\[ X_{n+1} = A_{n+1} X_n + u_{n+1}, \]

(9)

\[ X_{\theta,n+1} = A_{\theta,n+1} X_n + A_{n+1} X_{\theta,n} + u_{\theta,n+1}, \]

(10)

where \(A_{\theta,n+1}\) and \(u_{\theta,n+1}\) denotes the derivatives of \(A_{n+1}(\theta)\) and \(u_{n+1}(\theta)\) with respect to \(\theta\), respectively. Thus the state transition matrix will be

\[ \tilde{A}_n = \begin{pmatrix} A_n & 0 \\ A_{\theta,n} & A_n \end{pmatrix}. \]

We may then ask if \(\lambda(A) < 0\) implies \(\lambda(\tilde{A}) < 0\), and if an approximation similar to (6) is possible for the solutions of the extended system (9). A positive answer to this problem has been given in Mikosch and Straumann [15] under the condition that the sequence \((A_n)\) is independent and identically distributed (i.i.d.), and for some \(s > 0\)

\[ E \| \tilde{A}_1 \|^s < +\infty. \]

(13)

The proof of [15] follows a route completely different from what will be developed in the present paper, and apparently cannot be generalized to cover the general case to be discussed here.

The main result of the present paper is a positive answer to this problem in a general setting under significantly weaker conditions.

**Theorem 1.1.** Let

\[ A_n = \begin{pmatrix} A_n^1 & 0 \\ B_n & A_n^2 \end{pmatrix} \]

be a stationary, ergodic sequence of \((p+q) \times (p+q)\) matrices, satisfying (2), with \(A_n^1\) and \(A_n^2\) being square matrices. Then

\[ \lambda(A) = \max(\lambda(A^1), \lambda(A^2)). \]

Now consider a linear stochastic system

\[ X_{n+1}^1 = A_{n+1}^1 X_{n+1}^1 + u_{n+1} \]

(11)

\[ X_{n+1}^2 = A_{n+1}^2 X_{n+1}^2 + B_n X_{n+1}^1 + u_{n+1}^2, \]

(12)

where \(X_n = (X_n^1, X_n^2) \in \mathbb{R}^{p+q}\), \(u_n = (u_n^1, u_n^2)\), and \((A_n, u_n)\) is a jointly strictly stationary sequence. Assume that \(u_n\) satisfies condition (2). Then solving (11) with any initial condition \(X_0\) forward in time we get that for any \(\varepsilon > 0\)

\[ X_n - X_n^* = O(e^{(\lambda+\varepsilon)n}) \quad \text{w.p.1.} \]

(13)

An application and motivating example for the above theorem consider the problem of identifying the system parameters of a GARCH process. We proceed in a standard manner: for a fixed tentative value of the system parameters, say \(\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^\top\) we invert the system, using zero initial conditions, to get the assumed values \(\sigma_n^2(\theta)\) and \(\epsilon_n(\theta)\). Then the (conditional) quasi-log-likelihood function, modulo a constant, is

\[ L_N(\theta) = \sum_{n=1}^{N} -\frac{1}{2} \left( \log \sigma_n^2(\theta) + \frac{y_n^2}{2 \sigma_n^2(\theta)} \right). \]

To compute its gradient we need to differentiate \(\sigma_n^2(\theta)\). Using a state-space representation we are thus lead to a situation described in general above.

For a second application consider again the random linear stochastic system (1) satisfying \(\lambda(A) < 0\). We have seen that
condition (4) is sufficient for the existence of a stationary and causal solution. We give an example, using Theorem 1.1, which shows that this condition might be relaxed. Consider serially coupled input–output systems given by the equations, with \(-\infty < n < \infty\),

\[
X_n^1 = A_1^1 X_{n-1}^1 + u_n^1 \tag{14}
\]

\[
X_n^2 = A_2^2 X_{n-1}^2 + B_n X_{n-1}^1, \tag{15}
\]

where \(u_n^1 \in \mathbb{R}^p\). Note that there is no exogenous input in (15).

Defining

\[
u_n^2 = B_n X_{n-1}^1, \tag{16}
\]

we can write (15) as

\[
X_n^2 = A_2^2 X_{n-1}^2 + u_n^2. \tag{17}
\]

Note, however, that the validity of (4) for \(u_n^2\) cannot be guaranteed.

In spite of this (17) is well defined under the conditions of the following theorem, a direct consequence of Theorem 1.1.

**Corollary 1.1.** Let \((A_n, u_n^1)\) be a stationary, ergodic sequence, jointly satisfying (2). Assume that \(\lambda(A_1) < 0, \lambda(A_2) < 0\). Let \((X_n^1)\) be the unique strictly stationary, causal solution of (14).

Then the linear stochastic system (17) has a strictly stationary, causal solution given by

\[
X_n^2 = B_n X_{n-1}^1 + \sum_{k=1}^{\infty} A_2^2 A_{n-k-1}^{2n-k+1} u_{n-k}^2, \tag{18}
\]

where the right-hand side converges almost surely.

Consider again the linear stochastic system given by Eq. (1). Now assume that

\[
E|u_n|^q < +\infty \quad \text{for some } q \geq 1,
\]

where \(| \cdot |\) denotes the Euclidean norm in \(\mathbb{R}^p\). We may then ask under what condition will the infinite sum given in (5) converge in \(L_q\). This problem has been considered in [10] in a Markov chain setting. Assuming that \((A_n)\) and \((u_n)\) are i.i.d. sequences, which are also independent of each other, they have shown that for an even \(q\) a necessary and sufficient condition for \(L_q\)-convergence in (5) is that \(\rho(E(A_1^{\otimes q})) < 1\). Here \(\rho(.)\) is the spectral norm, and \(A^{\otimes q}\) denotes the \(q\)th Kronecker power of \(A\). A simpler proof for sufficiency under slightly weaker conditions has been given in [13]. Theorem 1.1 can be extended in the context of \(L_q\)-stability as follows, see [13]:

**Theorem 1.2.** Let

\[
A = \begin{pmatrix} A_1 & 0 \\ B & A_2 \end{pmatrix}
\]

be a random \((n_1 + n_2) \times (n_1 + n_2)\) matrix in \(L_q(\Omega, \mathcal{F}, P)\), with \(A_1\) and \(A_2\) being square matrices. Then for any positive integer, even or odd, we have

\[
\rho(E(A^{\otimes q})) = \max\{\rho(E(A_1^{\otimes q})); \rho(E(A_2^{\otimes q}))\}. \tag{19}
\]

**2. Proof of the main theorem**

It can be easily seen that the inequality \(\lambda(A) \geq \max(\lambda(A_1), \lambda(A_2))\) holds trivially. We will show that \(\lambda(A) \leq \min(\lambda(A_1), \lambda(A_2))\) holds as well. It is easy to see that it is sufficient to show that

\[
\lambda(A_1) < 0, \quad \lambda(A_2) < 0 \quad \text{implies } \lambda(A) < 0. \tag{20}
\]

Indeed, for any constant \(c > 0\) setting \(cA = (cA_n)\) obviously

\[
\lambda(cA) = \lambda(A) + \log c.
\]

Let \(c\) be such that \(\lambda(cA_1) < 0\) and \(\lambda(cA_2) < 0\). Then by (20) it follows that \(\lambda(cA) < 0\). In other words, with \(\gamma = \log c\), we have

\[
\lambda(A_1) < \gamma, \quad \lambda(A_2) < \gamma \quad \text{implies } \lambda(A) < \gamma,
\]

from which the claim follows. To prove (20) we use the following result which is part of Oseledec’s theorem, see [16, 17]. In the theorem below \(|x|\) stands for the Euclidean norm of a vector \(x \in \mathbb{R}^p\).

**Theorem 2.1.** Let \(A = (A_n)\) be a stationary, ergodic sequence of \(p \times p\) matrices satisfying (2). Then there are random subspaces \(S(\omega) \subset \mathbb{R}^p\) of fixed dimension, say, \(s < p\), such that \(S\) is a measurable function from \((\Omega, \mathcal{F})\) to the Grassmannian manifold of \(s\)-dimensional subspaces of \(\mathbb{R}^p\), and there exists a null-set \(\Omega_0 \subset \Omega\) such that for \(\omega \notin \Omega_0\) and any \(x \in \mathbb{R}^p, x \notin S(\omega)\) we have

\[
\lim_{n \to \infty} \frac{1}{n} \log |A_n \ldots A_1 x| = \lambda(A). \tag{21}
\]

**Remark 2.1.** The original form of Oseledec’s theorem has been stated for non-singular random matrices, see [16], but an extension to possibly singular sequences has been given in [17]. Although there seems to be a gap in the arguments of [17], the proof of the above partial result is rigorous.

**Corollary 2.1.** Under the conditions of Theorem 2.1 we have

\[
\lambda(A) < 0 \quad \text{if and only if for Lebesgue almost all } x \in \mathbb{R}^p \text{ we have}
\]

\[
\limsup_{n \to \infty} \frac{1}{n} \log |A_n \ldots A_1 x| < 0 \quad \text{w.p.1.} \tag{21}
\]

We note in passing that, obviously, the hard part of the above result is to show that (21) implies \(\lambda(A) < 0\).

To prove (20) we apply Corollary 2.1. Thus it is sufficient to show that for Lebesgue almost all \(x \in \mathbb{R}^{p+q}\)

\[
\limsup_{n \to \infty} \frac{1}{n} \log |A_n \ldots A_1 x| < 0 \quad \text{w.p.1.} \tag{22}
\]

Writing

\[
x_n^1 = A_n x_{n-1}^1 \quad x_0^1 = x^1
\]

\[
x_n^2 = B_n x_{n-1}^1 + A_n^2 x_{n-1}^2 \quad x_0^2 = x^2
\]
and solving this recursion we get
\[ x_n^1 = A_n^1 \ldots A_1^1 x_0^1 \] \hspace{1cm} (23)
\[ x_n^2 = A_n^2 \ldots A_2^2 x_0^2 + \sum_{k=1}^{n} A_n^2 \ldots A_k^2 A_{k-1}^1 \ldots A_1^1 x_0^1. \] \hspace{1cm} (24)
To estimate \( x_n^1 \) (3) is applicable. Thus we get
\[ |A_n^1 \ldots A_1^1 x_0^1| \leq C_A(\omega) e^{\lambda(1+\varepsilon)|x_0^1|} \] \hspace{1cm} (25)
with \( \lambda = \lambda(A) \). The first term on the right-hand side of (24) is estimated similarly. To estimate the norm of the second term of (24) we first apply the triangle inequality, and then estimate the \( k \)th term. We first estimate \( \|A_{k-1}^1 \ldots A_1^1\| \) using the inequality (3) with \((k-1)\) replacing \( n \). To estimate \( \|B_k\| \) we use the following simple well-known lemma:

**Lemma 2.1.** Let \( (\eta_k), k \geq 1 \) be a sequence of identically distributed real-valued random variables such that \( \mathbb{E} \log^+ |\eta_k| < +\infty \). Then for any \( \varepsilon > 0 \) there exists a finite random variable \( C(\omega) \) such that for all \( k \geq 1 \)
\[ |\eta_k| \leq C(\omega)e^{\lambda k}. \]
In other words \( (\eta_k) \) is sub-exponential.

Thus we get that for any \( \varepsilon > 0 \) there exists a finite random variable \( C(\omega) \) such that for all \( k \geq 1 \)
\[ \|B_k\| \leq C(\omega)e^{\lambda k}. \]
Therefore the norm of the second term in (24) can be estimated from above by
\[ C_A e^{\lambda} \sum_{k=1}^{n} \|A_n^2 \ldots A_{k+1}^2\| e^{\lambda(1+2\varepsilon)|x_0^1|}, \] \hspace{1cm} (26)
and here \( \lambda < 0 \). To estimate \( \|A_n^2 \ldots A_{k+1}^2\| \) we prove a few simple auxiliary results:

**Lemma 2.2.** Let \( (\xi_n), n \geq 1 \) be a strictly stationary, ergodic process with \( \mathbb{E}\xi_n = 0 \). Then for all \( \varepsilon > 0 \) there exists a random constant \( c(\omega) \) such that w.p.1
\[ \xi_1 + \cdots + \xi_n \leq c(\omega) + n\varepsilon. \]
**Proof.** The proposition is a direct consequence of the strong law of large numbers, known as Birkhoff’s ergodic theorem, for strictly stationary, ergodic processes. ■

Similarly, reversing the time, we would get for a two-sided, strictly stationary, ergodic process that
\[ \xi_n + \cdots + \xi_{k+1} \leq c_n(\omega) + (n-k)\varepsilon. \]
Unfortunately, the constant \( c_n(\omega) \) depends on \( n \). To get an upper bound in which the constant does not depend on \( n \), write
\[ \xi_n + \cdots + \xi_{k+1} = (\xi_1 + \cdots + \xi_n) - (\xi_1 + \cdots + \xi_k) \leq 2 \sup_{1 \leq k \leq n} (\xi_1 + \cdots + \xi_k) \leq 2(\xi_0 + n\varepsilon). \]
Repeating this argument with \( \varepsilon/2 \) replacing \( \varepsilon \) we get the following lemma:

**Lemma 2.3.** Let \( (\xi_n), n \geq 1 \) be a strictly stationary, ergodic process with \( \mathbb{E}\xi_n = 0 \). Then for all \( k \geq 0 \) there exists a random constant \( c(\omega) \) such that for all \( 0 \leq k < n \) we have w.p.1
\[ \xi_n + \cdots + \xi_{k+1} \leq c(\omega) + n\varepsilon. \]

In the case when \( \mathbb{E}\xi_n \neq 0 \) we have the following result:

**Lemma 2.4.** Let \( (\xi_n), n \geq 1 \) be a strictly stationary, ergodic process such that \( \mathbb{E}\xi_n \) exists and \( \mathbb{E}\xi_n < \infty \). Then for all finite \( \mu \geq \mathbb{E}\xi_n \), and for all \( \varepsilon > 0 \) there exists a random constant \( c(\omega) \), such that for all \( 0 \leq k < n \) we have w.p.1
\[ \xi_n + \cdots + \xi_{k+1} \leq c(\omega) + (n-k)\mu + n\varepsilon. \]

**Proof.** If \( \mathbb{E}\xi_n \) is finite, then, applying Lemma 2.3 for the random variables \( \xi_n = \xi_n - \mathbb{E}\xi_n \), the claim follows for \( \mu = \mathbb{E}\xi_n \), thus, by monotonicity, it also follows for larger \( \mu \)’s. If \( \mathbb{E}\xi_n = -\infty \), then first define \( \xi_n^\prime = \xi_n \vee (-K) \), where \( K \) is large enough to ensure that \( \mathbb{E}\xi_n^\prime \leq \mu \). Then the first part is applicable to get
\[ \xi_n^\prime + \cdots + \xi_{k+1} \leq c(\omega) + (n-k)\mu + n\varepsilon \quad \text{w.p.1}. \]
Since \( \xi_n^\prime \geq \xi_n \), we get the claim. ■

We continue the proof of Theorem 1.1 by estimating \( \|A_n^2 \ldots A_{k+2}\| \) first in the scalar case, i.e., when \( A_n^2 \) is a scalar, say \( |A_n^2| = a_n \). Then obviously \( \lambda^2 = \lambda(A^2) = \mathbb{E}\log a_n \).

**Lemma 2.5.** Let \( (a_n), n \geq 1 \) be a non-negative, strictly stationary, ergodic process such that \( \lambda = \mathbb{E}\log a_k \), and \( \lambda < \infty \). Then for all finite \( \mu \geq \lambda \), and for all \( \varepsilon > 0 \) there exists a random variable \( C(\omega) \), such that for all \( 0 \leq k < n \)
\[ a_n \cdots a_{k+1} \leq C(\omega) e^{(n-k)\mu+\varepsilon} \quad \text{w.p.1}. \]
A key point in the above statement is that \( C(\omega) \) is independent of \( n \).

**Proof.** Writing \( \xi_k = \log a_k \), applying Lemma 2.4, then exponentiating the resulting inequality, we get the claim. ■

Using this lemma with \( \lambda = \lambda(A^2) = \lambda^2 \) in (26), substituting the resulting upper bound into (24), and using the arguments following (24), we get after some simplifications that for \( \mu = \max(\lambda_1, \lambda_2) < 0 \), and for any \( \varepsilon \) there exists a random variable \( C'(\omega) \), such that for all \( n \geq 1 \) we have
\[ |x_n^2| \leq C'(\omega) e^{n(\mu+\varepsilon)}, \]
and thus (22) follows.

To estimate \( \|A_n^2 \ldots A_{k+2}\| \) in the general case we need a simple observation that states that the subadditive process \( \log |A_n^2 \ldots A_{k+2}| \) can be majorated by a scalar-valued additive process modulo negligible error with growth rate arbitrarily close to \( \lambda^2 = \lambda(A^2) < 0 \).
Lemma 2.6. Let \( A = (A_n) \) be a strictly stationary, ergodic sequence of \( p \times p \) random matrices such that \( \mathbb{E} \log^+ \| A_n \| < +\infty \). Then for any \( \varepsilon > 0 \) there exists a scalar-valued, stationary and ergodic process \( (\xi_n) \), and a finite random variable \( c(\omega) \), such that \( \mathbb{E} \xi_n < \lambda(A) + \varepsilon \), and for any \( 0 \leq k < n \)
\[
\log \| A_n \ldots A_{k+1} \| \leq c(\omega) + \xi_n + \ldots + \xi_{k+1} + n\varepsilon
\]
with probability 1.

It follows that, with \( a_k = e^{\xi_k} \) we have
\[
\| A_n \ldots A_{k+1} \| \leq C(\omega)a_n \ldots a_{k+1}e^{\varepsilon k},
\]
and here \( \mathbb{E} \log a_n < \lambda(A) + \varepsilon \).

The proof is essentially given in [11], and will be restyled and given in the Appendix. Applying the above result for \( A = A^2 = (A_n^2) \) in (24), the proof of Theorem 1.1 can be completed as in the scalar case.

3. Discussion

In this section we present a few remarks to highlight the delicacy of the details of the proof of the main result.

Remark 3.1. To estimate \( \| A_n \ldots A_{k+1} \| \), see Lemma 2.6, an alternative, direct approach would be to use the Furstenberg–Kesten theorem starting at time \( k \), and using the estimate that for all fixed \( \varepsilon > 0 \) we have
\[
\| A_n \ldots A_{k+1} \| \leq C_{k+1}(\omega) e^{(n-k)(\lambda+\varepsilon)},
\]
where \( \lambda = \lambda(A) \). Recall that \( C_{k+1}(\omega) \) can be defined as \( C_{k+1}(\omega) = e^{\zeta_{k+1}(\omega)} \), where
\[
\zeta_{k+1}(\omega) = \sup_{n \geq k+1} \left( \log \| A_n \ldots A_{k+1} \| - (n-k)(\lambda+\varepsilon) \right).
\]
Using a representation of \( (A_n) \) via a measure-preserving shift on \( \Omega \) it is easily seen that \( C_{k+1} \) can be assumed to be a stationary sequence. To control the effect of \( (C_{k+1}) \) in (24) we would need to show that \( C_{k+1} \) is sub-exponential. One way to show this would be to show that
\[
\mathbb{E} \log^+ C_{k+1}(\omega) < +\infty.
\] (28)

Unfortunately, this inequality is not true in general. Indeed, consider a scalar-valued i.i.d. process \( A_n = a_n \). Then \( \lambda = \mathbb{E} \log a_n \), and for fixed \( \varepsilon > 0 \) we have
\[
\log C_{k+1}(\omega) = \sup_{n \geq k+1} \sum_{j=k+1}^{n} (\log a_j - (\lambda + \varepsilon)).
\]
According to the Kiefer–Wolfowitz theorem (see [7], Chapter 10.4, Corollary 3) (28) holds if and only if
\[
\mathbb{E}(\log^+ a_j)^2 < +\infty.
\]

The same remark applies if we estimate \( \| A_n \ldots A_{k+1} \| \) using the Furstenberg–Kesten theorem backwards in time. Then we get for fixed \( n \), for all fixed \( \varepsilon > 0 \), and for all \( k \leq n \)
\[
\| A_n \ldots A_{k+1} \| \leq C_n(\omega)e^{(n-k)(\lambda+\varepsilon)},
\]
and here the growth rate of \( C_n(\omega) \) is, in general, not under control.

Remark 3.2. Consider Eq. (1), satisfying the conditions of Section 1, except \( \lambda(A) < 0 \). A possible route to establish \( \lambda(A) < 0 \) would be to establish the existence of a causal stationary solution directly, and then use the following deep result of [6].

Theorem 3.1. Consider the linear stochastic system given by the state-space equation (1), where \( X_n \in \mathbb{R}^p \), \( (A_n, u_n) \) is a jointly strictly stationary sequence of random matrices of size \( p \times (p+1) \), jointly satisfying condition (2). Let us assume that the sequence \( (A_n, u_n) \) is controllable in the sense that there is no proper subspace \( V \subset \mathbb{R}^p \), such that
\[
A_0 V + u_0 \subset V \quad \text{w.p.1}.
\]
Then if (1) has a strictly stationary causal solution \( (X_n) \), then \( \lambda(A) < 0 \).

Unfortunately, the direct verification of the existence of a causal stationary solution does not seem to be easy.

Remark 3.3. We note in passing, that an equally non-trivial deterministic version of the above result, with \( A_n \) taking its values from a given set of matrices \( A \), has been given in [12]. For more on this subject see [1,8].

Remark 3.4. The existence of the derivative process \( (X_{\theta,n}) \) in an almost sure sense, and, as a byproduct, the existence of a strictly stationary causal solution of (9) has been proved by direct arguments in [3] in the case of GARCH processes using specific arguments. Applying Theorem 3.1, we could conclude, with some additional work, that the top-Lyapunov exponent of the matrix
\[
\tilde{A}_n = \begin{pmatrix} A_n & 0 \\ \theta A_0 \end{pmatrix},
\]
is negative. It is not clear if this line of proof can used in the general case.

Remark 3.5. For the sake of completeness we present the main tool that is used in the proof of [15]:

Lemma 3.1. Let \( A = (A_n) \) be an i.i.d. sequence of \( p \times p \) matrices with \( E\|A_n\|^s < \infty \) for some \( s > 0 \). Then for the associated top-Lyapunov exponent we have \( \lambda(A) < 0 \) if and only if there exist \( c > 0 \), \( \bar{s} > 0 \) and \( 0 < \rho < 1 \) such that
\[
E\|A_n \ldots A_1\|^s \leq c\rho^s \quad \text{for } n \geq 1.
\]

The proof is based on the observation that for any fixed \( l \geq 1 \) the map \( u \mapsto E|A_l \ldots A_1|^u \) has first derivative with respect to \( u \) equal to \( E \log \| A_l \ldots A_1 \| \) at \( u = 0 \).

Acknowledgement

The authors acknowledge the support of the National Research Foundation of Hungary (OTKA) under Grant no. T 047193.
Appendix

Proof of Lemma 2.6. We follow the proof of [11]. For a fixed $\varepsilon > 0$ take an $l$ such that
\[ \frac{1}{l} \log \| A_l \| < \lambda + \varepsilon. \]
For any $0 \leq l < n$ and $0 \leq r < l$ take a cover of the index set \{ $k + 1, \ldots, n$ \} by $l$-tuples of the form
\[ I^\prime_q = \{ ql + r, \ldots, ql + r + l - 1 \}, \]
and let
\[ q_0 = q_0(r) = \min(q : ql + r > k + 1) \]
\[ q_1 = q_1(r) = \max(q : ql + r \leq n). \]
Set $k = q_0l + r$ and $\bar{n} = ql + r - 1$. Then
\[
\| A_n \ldots A_{k+1} \| \leq \| A_n \ldots A_{\bar{n}+1} \|
\]
\[
\times \| A_{\bar{n}+1} \ldots A_{\bar{n}+r-1} \ldots A_{\bar{n}+r-1} \ldots A_{\bar{n}+r-1} \ldots A_{k+1} \|, \]
thus $\log \| A_n \ldots A_{k+1} \|$ can be estimated from above by
\[
\log \| A_n \ldots A_{\bar{n}+1} \| + \sum_{q=q_0}^{q_1-1} \log \| A_{ql+r+l-1} \ldots A_{ql+r} \| + \log \| A_{\bar{n}+1} \ldots A_{k+1} \|. \tag{29}
\]
By Lemma 2.1 we have that for any $\varepsilon' > 0$ there exists a random variable $c'(\omega)$ such that $\log \| A_1 \| \leq c'(\omega) + \varepsilon'k$. Thus, since $n - (\bar{n} + 1) \leq l$ we have
\[
\log \| A_n \ldots A_{\bar{n}+1} \| \leq \log \| A_n \| + \ldots + \log \| A_{\bar{n}+1} \| \leq l(c'(\omega) + \varepsilon'n).
\]
A similar inequality holds for $\log \| A_{\bar{n}+1} \ldots A_{k+1} \|$. Now define $\xi_i = \log \| A_{i+l-1} \ldots A_i \|$, $\eta^\prime_q = \log \| A_{ql+r+l-1} \ldots A_{ql+r} \|$. Note that $\eta^\prime_q = \xi_{ql+r}$. Then the middle term in (29) can be written as
\[
\sum_{q=q_0}^{q_1-1} \eta^\prime_q = \sum_{q=q_0}^{q_1-1} \xi_{ql+r}.
\]
Letting $r$ run from 0 to $(l - 1)$ and averaging (29) over $r$ we get
\[
\log \| A_n \ldots A_{k+1} \| \leq \left( \frac{1}{l} \sum_{i=k+1}^{n-l} \xi_i \right) + 2l(c'(\omega) + \varepsilon'n)
\]
\[
\leq \left( \frac{1}{l} \sum_{i=k+1}^{n} \xi_i \right) + 3l(c'(\omega) + \varepsilon'(n + l)) \quad \text{a.s.},
\]
since $\xi_i \leq l(c'(\omega) + \varepsilon'(n + l))$ for any $i \leq n$. For fixed $l$ taking $\varepsilon'$ sufficiently small we can upper bound the last, residual term as $c''(\omega) + \varepsilon n$ for any prescribed $\varepsilon$.

By assumption $\frac{1}{l} E \xi_l \leq \lambda + \varepsilon$, and $(\xi_\omega)$ is strictly stationary and ergodic, hence, applying Lemma 2.4 we get that with some $c'(\omega)$ depending on $\varepsilon$ we have
\[
\log \| A_n \ldots A_{k+1} \| \leq \frac{1}{l} \sum_{i=k+1}^{n} \xi_i + (c''(\omega) + \varepsilon n)
\]
\[
\leq c(\omega) + (n - k)(\lambda + \varepsilon) + n\varepsilon + (c'(\omega) + \varepsilon n) \quad \text{a.s.}
\]
which implies the claim. ■

References