Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Stochastic Processes and their Applications 118 (2008) 1997–2013
stochastic processes and their applications
www.elsevier.com/locate/spa

Weakly dependent chains with infinite memory

Paul Doukhan[a,b], Olivier Wintenberger[b,*]

[a] *LS-CREST, Laboratoire de Statistique, Timbre J340, 3, avenue Pierre Larousse, 92240 Malakoff, France*
[b] *SAMOS-MATISSE (Statistique Appliquée et MOdélisation Stochastique), Centre d'Économie de la Sorbonne Université Paris 1 - Panthéon-Sorbonne, CNRS 90, rue de Tolbiac 75634-Paris Cedex 13, France*

Received 16 November 2006; received in revised form 8 December 2007; accepted 17 December 2007
Available online 27 December 2007

Abstract

We prove the existence of a weakly dependent strictly stationary solution of the equation $X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \ldots; \xi_t)$ called a *chain with infinite memory*. Here the *innovations* ξ_t constitute an independent and identically distributed sequence of random variables. The function F takes values in some Banach space and satisfies a Lipschitz-type condition. We also study the interplay between the existence of moments, the rate of decay of the Lipschitz coefficients of the function F and the weak dependence properties. From these weak dependence properties, we derive strong laws of large number, a central limit theorem and a strong invariance principle.
© 2007 Elsevier B.V. All rights reserved.

MSC: primary 62M10; secondary 91B62; 60K35; 60K99; 60F05; 60F99

Keywords: Time series; Weak dependence; Central limit theorems; Uniform laws of large numbers; Strong invariance principles

1. Introduction

Statistical inferences heavily rely on the underlying model. The same process may have different representations and it may belong to different classes of models. In this paper, we introduce a *chain with infinite memory* as the stationary solution of the equation

$$X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \ldots; \xi_t), \quad \text{a.s. for } t \in \mathbb{Z}, \tag{1.1}$$

* Corresponding author. Tel.: +33 (0)144078935; fax: +33 (0)144078922.
E-mail addresses: doukhan@ensae.fr (P. Doukhan), olivier.wintenberger@univ-paris1.fr (O. Wintenberger).

0304-4149/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2007.12.004

where F takes values in a Banach space. For details, see Section 3.1. The dynamical behavior described by (1.1) corresponds to a large variety of time series models. Those models can be seen as natural extensions, either of linear models or of Markov models. In the sequel, the *innovations* ξ_t constitute an independent and identically distributed (iid) sequence. *Chains with infinite memory* can be represented as causal Bernoulli shifts $X_t = H(\xi_t, \xi_{t-1}, \ldots)$, and then conditions on H gave asymptotic results, see [31]. But several Bernoulli shifts, such as Volterra series, may not fit the parsimony criterion and the function H may be non-explicit. This is a drawback for statistical inferences in that context. Autoregressive representations are preferred in various applications, e.g. in finance, hydrodynamics, physics; see [11,25]. Kallenberg [23] stresses the fact that all the p-Markov processes are solutions of equations of the type

$$X_t = F(X_{t-1}, \ldots, X_{t-p}; \xi_t). \tag{1.2}$$

Bougerol [3] gave conditions of Lyapunov type for the existence of a stationary solution to *Stochastic Recurrence Equations* (SRE), which are particular cases of (1.2).

Approaches other than (1.1) to modeling processes which do not satisfy the Markov property already exist; the *Random Systems with Complete Connections* (RSCC; see [21]) and the *Variable Length Markov Chains* (VLMC; see [4]). Such models are widely used in the fields of particle systems or in DNA data analysis. These processes are defined through their conditional distributions. Their existence relies on assumptions on the conditional expectations, following the work of Dobrushin [10]. Notice that Berbee [2] obtained another existence condition for the cases where the state space is discrete; see also [5,18].

Dobrushin's condition implies strong mixing; see [12,21]. Mixing coefficients are useful for deriving asymptotic theorems for various functionals of a stationary sequence; see Rio [30]. However, major asymptotic results still hold under so-called weak dependence conditions; see Section 2.2, [8,13] and the recent monograph by [7]. The Central Limit Theorem (CLT) of Dedecker and Doukhan [6] holds if the $x^2 \ln(1 + x)$th moments of X_0 are finite and if the process is weakly dependent with geometric decay of the coefficients. Because weak dependence is less restrictive than mixing (see Andrews [1] for an example) this result extends the CLT for mixing sequences due to Rio [30]. The conditions for those CLT are expressed in terms of Orlicz functions that balance the moments of some order and the weak dependence conditions.

The existence of a stationary solution to (1.1) is proved in Section 3.2 under a specific Lipschitz-type assumption on F; see (3.1). Approximation by suitable Markov processes is the main tool for the proofs given in Section 5. This existence condition also yields finiteness of moments of some order in terms of Orlicz functions. We get bounds for the weak dependence coefficients of the solution to (1.1). We use these bounds to derive sufficient conditions on F in terms of Orlicz functions and in turn to prove a Strong Law of Large Numbers (SLLN), a CLT and a Strong Invariance Principle (SIP); see Section 3.3. We discuss the generality of our model in Section 4 comparing it with existing ones. But to begin with, we introduce some notation and we define useful tools such as weak dependence coefficients and Orlicz spaces.

2. Preliminaries

2.1. Notation

In the sequel, the iid innovations ξ_t for $t \in \mathbb{Z}$ take values in a measurable space $(E', \mathcal{A}')$. Let $\|\cdot\|$ denote the norm of a Banach space E. The space $E^{(\infty)}$ is the subset of $E^{\mathbb{N}}$ of finitely non-zero sequences $(x_k)_{k>0}$ such that there exists $N > 0$ with $x_k = 0$ for $k > N$. Let E be endowed with its Borel σ-algebra $\mathcal{A}$; then $E^{(\infty)}$ is considered together with its product σ-algebra $\mathcal{A}^{\otimes\mathbb{N}}$. The function F in (1.1) is assumed to be a measurable function from $E^{(\infty)} \times E'$ with values in E. Moreover $\|\cdot\|_m$ denotes the usual $\mathbb{L}^m$-norm, *i.e.*, $\|X\|_m^m = \mathbb{E}\|X\|^m$ for $m \geqslant 1$ for every E-valued random variable X. For $h : E \to \mathbb{R}$, we define $\|h\|_\infty = \sup_{x\in E} |h(x)|$ and

$$\mathrm{Lip}\,(h) = \sup_{x\neq y} \frac{|h(x) - h(y)|}{\|x - y\|}.$$

The space $\Lambda_1\,(E)$ is the set of functions $h : E \to \mathbb{R}$ such that $\mathrm{Lip}\,(h) \leqslant 1$.

2.2. *Weak dependence*

An appropriate notion of weak dependence for the model (1.1) was introduced in [8]. It is based on the concept of the coefficient τ defined below. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-subalgebra of $\mathcal{C}$ and X a random variable with values in E. Assume that $\|X\|_1 < \infty$ and define the coefficient τ as

$$\tau(\mathcal{M}, X) = \left\| \sup\left\{ \left| \int f(x)\mathbb{P}_{X|\mathcal{M}}(\mathrm{d}x) - \int f(x)\mathbb{P}_X(\mathrm{d}x) \right| \text{ with } f \in \Lambda_1\,(E) \right\} \right\|_1 .$$

An easy way to bound this coefficient is based on a coupling argument:

$$\tau(\mathcal{M}, X) \leqslant \|X - Y\|_1$$

for any Y with the same distribution as X and independent of $\mathcal{M}$; see [8]. Moreover, if the probability space $(\Omega, \mathcal{C}, \mathbb{P})$ is rich enough (we always assume so in the sequel) there exists an X^* such that $\tau(\mathcal{M}, X) = \|X - X^*\|_1$. Using the definition of τ, the dependence between the past of the sequence $(X_t)_{t\in\mathbb{Z}}$ and its future k-tuples may be assessed: Consider the norm $\|x - y\| = \|x_1 - y_1\| + \cdots + \|x_k - y_k\|$ on E^k, set $\mathcal{M}_p = \sigma(X_t, t \leqslant p)$ and define

$$\tau_k(r) = \max_{1\leqslant l\leqslant k} \frac{1}{l} \sup\{\tau(\mathcal{M}_p, (X_{j_1}, \ldots, X_{j_l})) \quad \text{with } p + r \leqslant j_1 < \cdots < j_l\},$$
$$\tau_\infty(r) = \sup_{k>0} \tau_k(r).$$

For the sake of simplicity, $\tau_\infty(r)$ is denoted by $\tau(r)$. Finally, the time series $(X_t)_{t\in\mathbb{Z}}$ is τ-*weakly dependent* when its coefficients $\tau(r)$ tend to 0 as r tends to infinity.

2.3. *Orlicz spaces*

Orlicz spaces are convenient generalizations of the classical $\mathbb{L}^m$-spaces; we refer the reader to [24] for the introduction and properties of such spaces. Let Φ be an Orlicz function, *i.e.*, defined on $\mathbb{R}^+$, convex, increasing and satisfying $\Phi(0) = 0$. For any random variable X with values in E, the norm $\|X\|_\Phi$ is defined by the equation

$$\|X\|_\Phi = \inf\left\{ u > 0 \text{ with } \mathbb{E}\left[\Phi\left(\frac{\|X\|}{u}\right)\right] \leqslant 1 \right\}.$$

The Orlicz space $\mathbb{L}^\Phi$ is given by

$$\mathbb{L}^\Phi = \{E\text{-valued random variables } X \text{ such that } \|X\|_\Phi < \infty\}.$$

It is a Banach space equipped with the norm $\|\cdot\|_\Phi$. For $m \geqslant 1$ and $\Phi(x) = x^m$, notice that $\mathbb{L}^\Phi$ is the usual $\mathbb{L}^m$-space. We restrict ourselves to Orlicz functions Φ satisfying the condition:

$$\text{For all } x, y \in \mathbb{R}^+, \quad \Phi(xy) \leqslant \Phi(x)\Phi(y). \tag{2.1}$$

This class of Orlicz functions is sufficiently large. For instance, the functions $\Phi(x) = x^m$ and $\Phi(x) = x^m(1 + \ln(1 + x))^{m'}$ satisfy (2.1) for each $m \geqslant 1, m' \geqslant 0$. Moreover, if ϕ is any Orlicz function satisfying the Δ_2-condition (there exists $k > 0$ such that $\phi(2x) \leqslant k\phi(x)$) then $\Phi(x) = \sup_{u>0} \phi(xu)/\phi(u)$ is an Orlicz function satisfying (2.1). Various examples of Orlicz functions satisfying the Δ_2-condition are given in [24].

Later, in Theorem 3.2 we will need some transformations of Orlicz functions. Given such a function Φ, we define, for $q > 1$,

$$\widetilde{\Phi}_q(x) = \sup_{y>0}\{(xy)^{q-1} - \Phi(y)/y\}. \tag{2.2}$$

The transformations $\widetilde{\Phi}_q(x)$ have simple bounds for certain choices of Φ; see Lemma 5.1 for details. In particular, if $\Phi(x) = x^m$ for $m > q > 1$, then $\widetilde{\Phi}_q(x) \leqslant x^{(m-1)(q-1)/(m-q)}$. Another useful example is the one of $\Phi(x) = x^q(1 + \ln(1 + x))^{(1+b)(q-1)}$ and $\widetilde{\Phi}_q(x) \leqslant \exp((q-1)x^{1/(1+b)})x^{q-1}$ for any $q > 1$ and $b \geqslant 0$.

3. The results

3.1. Assumptions

The existence of a solution to (1.1) will be proved under a Lipschitz-type condition. We express it in terms of some Orlicz functions in order to be able to work with moments more general than power moments; see Theorem 3.1. These moments will be needed to establish the asymptotic results of Theorem 3.2.

Assume there exists an Orlicz function Φ such that for all x, y in $E^{(\infty)}$

$$\|F(x; \xi_0) - F(y; \xi_0)\|_\Phi \leqslant \sum_{j=1}^{\infty} a_j \|x_j - y_j\|, \tag{3.1}$$

where $(a_j)_{j \geqslant 1}$ is a sequence of non-negative real numbers such that

$$a = \sum_{j=1}^{\infty} a_j < 1 \quad \text{and} \tag{3.2}$$

$$\mu_\Phi = \|F(0, 0, \ldots; \xi_0)\|_\Phi < \infty. \tag{3.3}$$

The Lipschitz property of F and the moment assumption (3.3) induce that $\|F(c; \xi_0)\|_\Phi < \infty$ for any constant $c \in E^{(\infty)}$. We choose $c = (0, 0, \ldots)$ in condition (3.3) for convenience.

3.2. Existence, moments and weak dependence

The following theorem settles the existence of a solution to (1.1). It also states that the Φth moment of this solution is finite.

Theorem 3.1. *Assume that conditions (3.2) and (3.3) hold for some Orlicz function Φ satisfying (2.1). Then there exists a τ-weakly dependent stationary solution $(X_t)_{t\in\mathbb{Z}}$ of (1.1) such that $\|X_0\|_\Phi < \infty$ and*

$$\tau(r) \leqslant 2\frac{\mu_1}{1-a}\inf_{1\leqslant p\leqslant r}\left(a^{r/p}+\frac{1}{1-a}\sum_{k=p+1}^{\infty}a_k\right)\to 0 \quad \text{as } r\to\infty.$$

The proof of the existence of a solution to (1.1) is given in Section 5.3 expressing it as the limit of the p-Markov processes defined in (1.2). The weak dependence properties are proved in Section 5.4.

Remark 3.1. We also prove in Section 5 that there exists some measurable function H such that $X_t = H(\xi_t, \xi_{t-1}, \ldots)$. This means that the process $(X_t)_{t\in\mathbb{Z}}$ can be represented as a causal Bernoulli shift. For those processes, conditions (3.2) and (3.3) together imply the Dobrushin uniqueness condition; see [10]. Thus $(X_t)_{t\in\mathbb{Z}}$ is the unique causal Bernoulli shift solution to (1.1). Moreover, as a causal Bernoulli shift, the solution $(X_t)_{t\in\mathbb{Z}}$ is automatically an ergodic process. Under the conditions of Theorem 3.1, the solution to (1.1) has finite Φth moment. From Lemma 5.3, $(X_t)_{t\in\mathbb{Z}}$ has also finite first-order moments. The ergodic theorem yields the SLLN for any *chain with infinite memory* under the assumptions of Theorem 3.1.

Corollary 3.1. *Under the assumptions of Theorem 3.1, there exists a τ-weakly dependent stationary solution $(X_t)_{t\in\mathbb{Z}}$ to (1.2) such that $\|X_0\|_\Phi < \infty$ and $\tau(r) \leqslant 2\mu_1(1-a)^{-1}a^{r/p}$ for $r \geqslant p$.*

Dedecker and Prieur [8] proved the existence of a solution to (1.2). They stated that there exists $0 < \rho < 1$ and $C > 0$ such that $\tau(r) \leqslant C\rho^r$. Applying Corollary 3.1, we get the bound $\rho \leqslant a^{1/p}$. The bounds of the weak dependence coefficients in Theorem 3.1 come from an approximation with Markov chains of order p and from the result of Corollary 3.1.

In Theorem 3.1, the τ-weak dependence property is linked to the choice of the parameter p and then to the rate of decay of the Lipschitz coefficients a_j. For example, if $a_j \leqslant ce^{-\beta j}$, we choose p as the largest integer smaller than $\sqrt{-\ln(a)r/\beta}$ to derive the bound $\tau(r) \leqslant Ce^{-\sqrt{-\ln(a)\beta r}}$ for some suitable constant $C > 0$. If $a_j \leqslant cj^{-\beta}$, we choose the largest integer p such that $p\ln p(1-\beta)/\ln a \leqslant r$. Then there exists $C > 0$ such that $\tau(r) \leqslant Cp^{1-\beta}$. Notice that $\ln r$ is smaller than $\ln p + \ln\ln p$ up to a constant and that $\ln r/r$ is proportional to $1/p(1+\ln\ln p/\ln p)$ and then equivalent to $1/p$ as p tends to infinity with r. From these equivalences, we achieve thus that there exists $C > 0$ such that $\tau(r) \leqslant C\,(\ln r/r)^{\beta-1}$.

A result similar to the one of Theorem 3.1 was obtained for discrete state space models (such as RSCC) in [20]. They gave bounds for the mixing coefficients under conditions on the marginal distributions of the *innovations*. The bound in [20], Theorem 2.1.5 on page 42, is similar to the one for $\tau(r)$ in Theorem 3.1. In a sense we extend their result: Here the *innovations* are not supposed to be absolutely continuous and our approach can be applied to discrete state space processes as well; see the example of the *Galton–Watson process with immigration* in Section 4.

Bougerol gives in [3] a recursive approximation of the stationary measure in the Markovian case. In Proposition 3.1 below we generalize this result to the infinite memory case. Let $\phi_k : E^{k-1}\times E' \to E$ be the random function defined as $x \mapsto F(x, c; \xi_k)$, for each $k \geqslant 2$ and some fixed sequence $c = (c_1, c_2, \ldots) \in E^{(\infty)}$. Write $\widetilde{X}_1 = \phi_1 = \phi(c; \xi_0)$ and define recursively

$$\widetilde{X}_n = \phi_n(\widetilde{X}_{n-1}, \ldots, \widetilde{X}_1).$$

Proposition 3.1. *Assume that conditions (3.2) and (3.3) hold for Φ satisfying (2.1). If $(X_t)_{t\in\mathbb{Z}}$ is the solution to (1.1) then*

$$\|\widetilde{X}_r - X_r\|_\Phi \leqslant \frac{\|X_0\|_\Phi + \bar{c}}{1-a} \inf_{1\leqslant p\leqslant r}\left(a^{r/p} + \frac{1}{1-a}\sum_{k=p+1}^{\infty} a_k\right) \to 0 \quad \text{as } r\to\infty,$$

where $\bar{c}$ is a constant such that $\|c_i\| \leqslant \bar{c}$ for all $i \geqslant 1$.

The proof of this proposition is given in Section 5.6.

3.3. Asymptotic results

In this section, $E = \mathbb{R}$. We give an appropriate condition on F (see (3.4)) that leads to versions of the results of Dedecker and Doukhan [6] and Dedecker and Prieur [8] obtained under weak dependence.

Theorem 3.2. *Assume that conditions (3.2) and (3.3) hold for some Orlicz function Φ satisfying (2.1) and assume there exists $c_0 > 0$ such that*

$$\sum_{k\geqslant 1} a_k \widetilde{\Phi}_q(c_0 k) < \infty \quad \text{if there exist } p \geqslant 1 \text{ such that } \sum_{j>p} a_j = 0, \tag{3.4a}$$

$$\sum_{k\geqslant 1} a_k \widetilde{\Phi}_q\left(-c_0 k \ln\left(\sum_{j\geqslant k} a_j\right)\right) < \infty \quad \text{otherwise}, \tag{3.4b}$$

where $\widetilde{\Phi}_q$ is defined in (2.2). The following relations hold:

SLLN: *If $q \in]1,2[$ then $n^{-1/q}\sum_{i=1}^n (X_i - \mathbb{E}X_0) \to_{n\to\infty} 0$, a.s.*

CLT: *If $q = 2$, then $\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(X_i - \mathbb{E}X_0) \xrightarrow{D[0,1]} \sigma W(t)$ as $n\to\infty$ where $\sigma^2 = \sum_{i=-\infty}^{\infty}\mathrm{Cov}(X_0, X_i)$ is finite and $W(t)$ is the standard Wiener process.*

SIP: *If $q = 2$ and if the underlying probability space is rich enough then there exist independent $\mathcal{N}(0,\sigma^2)$-distributed random variables $(Y_i)_{i\geqslant 1}$ such that*

$$\sum_{i=1}^n (X_i - Y_i) = o(\sqrt{n \ln\ln n}) \quad \text{a.s.}$$

The proof of this theorem is given in Section 5.5.

Note that $x^2\ln(1+x)$th moments are necessary to get the CLT for weakly dependent processes. See [15] for an example of processes, solutions of (1.2) for $p = 1$, that do not satisfy the CLT under conditions (3.2) and (3.3) for $\Phi(x) = x^2$. Note also that approximations by martingale difference as in [28] or projective criterion as in [27] give the CLT under weaker assumptions for some of the examples treated in Section 4.

Condition (3.4a) is relevant for the Markov solution $(X_t)_{t\in\mathbb{Z}}$ to (1.2), *i.e.*, when $\sum_{j>p} a_j = 0$. For the other cases, we rewrite assumption (3.4b) for various rates of decay of the Lipschitz coefficients a_j. Let a, b, c be some positive real numbers; then

$$\text{If } a_k \leqslant ck^{-a}, \quad \sum_{k\geqslant 1} a_k \widetilde{\Phi}_q(c_0 k \ln k) < \infty \quad \text{for some } c_0 > 0. \tag{3.4b$'$}$$

$$\text{If } a_k \leqslant c\exp(-ak^b), \quad \sum_{k\geqslant 1} a_k \widetilde{\Phi}_q\left(c_0 k^{1+b}\right) < \infty \quad \text{for some } c_0 > 0. \tag{3.4b$''$}$$

For instance, condition (3.4b′) holds if $\Phi(x) = x^m$ for $m > q$ and $a > 1 + (q-1)(m-1)(m-q)^{-1}$. Condition (3.4b″) holds for $\Phi(x) = x^q(1+\ln(1+x))^{(1+b)(q-1)}$. Applying Theorem 3.2, the CLT and the SIP hold for sub-geometric rates of decay of the Lipschitz coefficients as in (3.4b″) under a moment condition of order $x^2(1+\ln(1+x))^{1+b}$.

4. Examples

In this section, we present some examples with $E = \mathbb{R}^d$ and $d \geqslant 1$. We consider the finite memory case as well as an infinite memory extension of *Stochastic Recurrence Equations* (SRE). In particular, we consider the example of the *Galton–Watson process with immigration* which satisfies the conditions of our results, but it is not a SRE in the sense of [3].

4.1. Markov models

SRE. We consider an iid process $(\phi_t)_{t\in\mathbb{Z}}$ of random Lipschitz maps with $\|\phi_t(x_1) - \phi_t(y_1)\| \leqslant L(\phi)\|x_1 - y_1\|$ a.s. for all $x_1, y_1 \in E$ and $t \in \mathbb{Z}$. Moreover let $\phi_t(x)$ be measurable for every fixed $x \in E$ and $t \in \mathbb{Z}$. If a stochastic process $(X_t)_{t\in\mathbb{Z}}$ with values in E satisfies the equation

$$X_{t+1} = \phi_t(X_t) \quad \text{a.s., for all } t \in \mathbb{Z},$$

we say that $(X_t)_{t\in\mathbb{Z}}$ obeys the SRE associated with $(\phi_t)_{t\in\mathbb{Z}}$. We write this equation as in (1.2) setting $\xi_t = \phi_t$ for $t \in \mathbb{Z}$, and $F(x,z) = z(x)$ for $x \in E$ and $z \in E'$, the space of Lipschitz random functions. In this case, conditions (3.2) and (3.3) become

$$\|L(\phi)\|_\Phi < 1 \quad \text{and} \quad \|\phi_0(0)\|_\Phi < \infty.$$

Weaker conditions related to a Lyapunov exponent for the existence of an a.s. solution to a SRE are obtained in [3]. However, that result does not yield the existence of moments or asymptotic results such as those in Theorem 3.2. We also mention the survey article by [9] for an overview and nice applications of SREs.

Non-linear autoregressive models. Here we consider a solution to (1.1), where $E' = E$ and F admits the representation

$$F(x_1, \ldots, x_p; s) = R(x_1, \ldots, x_p) + s.$$

Condition (3.2) becomes

$$\|R(y_1, \ldots, y_p) - R(x_1, \ldots, x_p)\| \leqslant \sum_{j=1}^{p} a_j \|x_j - y_j\| \quad \text{with } \sum_{j=1}^{p} a_j < 1,$$

and condition (3.3) coincides with $\|\xi_0\|_\Phi < \infty$. Results similar to those in Theorem 3.2 are obtained by different methods in [17].

Galton–Watson processes with immigration. If $E = \mathbb{R}$, a *Galton–Watson process with immigration* is given as a stationary solution of the equation

$$X_t = \begin{cases} \displaystyle\sum_{i=1}^{X_{t-1}} \zeta_{t,i} + \zeta_t, & \text{if } X_{t-1} > 0, \\ \zeta_t & \text{if } X_{t-1} = 0. \end{cases} \tag{4.1}$$

Here $(\zeta_{t,i})_{t\in\mathbb{Z}, i>0}$, $(\zeta_t)_{t\in\mathbb{Z}}$ are independent iid families of integer-valued random variables and $E' = \mathbb{N}^{\mathbb{N}}$ is equipped with the product measure. We can write $X_t = F(X_{t-1}, \xi_t)$ with

$F(x,(u_i)_{i\geqslant 0}) = u_0 + \sum_{i=1}^{x} u_i$ if $x > 0$ and $F(0,(u_i)_{i\geqslant 0}) = 0$ for any $(u_i)_{i\geqslant 0}$. If $y_1 > x_1 > 0$ then $F(x,(u_i)_{i\geqslant 0}) - F(x,(u_i)_{i\geqslant 0}) = \sum_{i=x_1}^{y_1} u_i$ and thus

$$\|F(x_1,\xi_0) - F(y_1,\xi_0)\|_\Phi = \left\| \sum_{i=x_1}^{y_1} \zeta_{0,i} \right\|_\Phi = |y_1 - x_1| \|\zeta_{0,0}\|_\Phi.$$

Assumptions (3.2) and (3.3) hold as soon as $\|\zeta_{0,0}\|_\Phi < 1$. This model is not a SRE if $\zeta_{0,0}$ is not finitely supported; thus we are not under the conditions of [3]. Other non-SRE examples which can be treated by our approach are given in [26].

4.2. SRE with infinite memory

Infinite memory extensions of classical SRE are solutions of the equation

$$\begin{cases} X_t = \phi_t(X_{t-1}, X_{t-2}, \ldots) \quad \text{a.s.,} \\ \|\phi_t(x) - \phi_t(y)\| \leqslant \sum_{i=1}^{\infty} L_i(\phi)\|x_i - y_i\|, \quad \text{a.s.} \end{cases}$$

for all $x = (x_i)_{i\geqslant 1}, y = (y_i)_{i\geqslant 1} \in E^{(\infty)}$. Here $(\phi_t)_{t\in\mathbb{Z}}$ is an iid process of random Lipschitz maps. If $\sum_{i\geqslant 1} \|L_i(\phi)\|_\Phi < 1$ then conditions (3.2) and (3.3) are satisfied. Some examples with this representation follow.

Non-linear ARCH (∞) *models.* Here $(X_t)_{t\in\mathbb{Z}}$ is the stationary solution of the equation

$$X_t = \xi_t \left(\alpha + \sum_{j=1}^{\infty} \alpha_j(X_{t-j}) \right),$$

where ξ_t is a $d \times k$ matrix, $E' = \mathcal{M}_{k,d}(\mathbb{R})$, $\alpha \in \mathbb{R}^k$ and $\alpha_j : E \to \mathbb{R}^k$ are Lipschitz functions. The LARCH(∞) model of [19,16] corresponds to the special case of linear functions $\alpha_j(x) = c_j x$ with $k \times d$ matrices c_j. Assumptions (3.2) and (3.3) hold as soon as $\|\xi_0\|_\Phi \sum_{j\geqslant 1} \mathrm{Lip}\,\alpha_j < 1$ and $\sum_{j\geqslant 1} \alpha_j(0) < \infty$.

Models with linear input. Let $f : \mathbb{R}^k \times E' \to E$ be measurable and satisfy $\|f(t,\xi_0) - f(s,\xi_0)\|_\Phi \leqslant L\,\|t-s\|$ for some finite constant $L > 0$. We consider

$$X_t = f(A_t, \xi_t), \qquad A_t = \sum_{j=1}^{\infty} c_j X_{t-j},$$

where c_j are $k \times d$ matrices. Relations (3.2) and (3.3) hold if $L \sum_{j\geqslant 1} \|c_j\| < 1$ and $f(0,\xi_0)\,\|_\Phi < \infty$. These models are used in statistical mechanics; see [22].

Affine models. Let us consider the special case of chains with infinite memory that can be written in a bilinear form

$$X_t = M_t \xi_t + f_t, \tag{4.2}$$

where $M_t = M(X_{t-1}, X_{t-2}, \ldots)$ and $f_t = f(X_{t-1}, X_{t-2}, \ldots)$ are both Lipschitz functions of the past values $X_{t-1}, X_{t-2}, X_{t-3}, \ldots$. Applying Theorem 3.1 under the condition

$$\|\xi_0\|_\Phi \sum_{i=1}^{\infty} \mathrm{Lip}\, M_i + \sum_{i=1}^{\infty} \mathrm{Lip}\, f_i < 1,$$

there exists a weakly dependent solution to (4.2). This class contains various time series models (such as ARCH, GARCH, ARMA, ARMA-GARCH, etc.). In the appendix we prove the existence of the joint densities of the solution to (4.2). This result and the weak dependence properties obtained in Theorem 3.1 are needed for achieving optimal rates of convergence of non-parametric estimators; see [29].

5. Proofs of the main results

After some preliminaries in Section 5.1, in Section 5.2 we construct a solution of the Markov model (1.2). We use it to approximate the solution to (1.1). The existence of a solution to (1.1), presented in Theorem 3.1, is obtained as $p \to \infty$ in Section 5.3. Its weak dependence properties are derived by coupling techniques in Section 5.4. Using weak dependence results of [6,8], we prove Theorem 3.2 in Section 5.5. Finally, we derive the proof of Proposition 3.1 in Section 5.6.

5.1. Preliminaries

We first present four useful lemmas. The first one aims at bounding the transformations $\tilde{\Phi}_q$ for $q > 1$; the other ones are used in the proof of the existence of a solution of (1.1).

Lemma 5.1. *Assume L is an increasing non-negative function on $[0, \infty]$ and write L^{-1} for the generalized inverse of L, i.e., $L^{-1}(x) = \inf\{y > 0$ with $L(y) \geqslant x\}$. If $\Phi(x) = x^q L(x)$, $x \geqslant 0$, for some $q > 1$ then*

$$\widetilde{\Phi}_q(x) \leqslant \left(x L^{-1}(x^{q-1})\right)^{q-1} \quad \textit{for all } x \geqslant 0.$$

Proof. From (2.2), we have $\widetilde{\Phi}_q(x) = \sup_{y>0}\{y^{q-1}(x^{q-1} - L(y))\}$. We restrict ourselves to $y \leqslant L^{-1}(x^{q-1})$; otherwise $y^{q-1}(x^{q-1} - L(y)) \leqslant 0$. Now notice that the first term of the product $y^{q-1}(x^{q-1} - L(y))$ is increasing and the second term always remains smaller than x^{q-1}. This proves the lemma. □

Lemma 5.2. *Assume that the Orlicz function Φ satisfies (2.1). Let ξ and ζ be independent random variables, z a measurable function and $Z = z(\xi, \zeta)$. We write $\mathbb{E}_\xi$ for the expectation with respect to the distribution of ξ. Define*

$$\|z(\xi, \zeta)\|_{\Phi,\xi} = \inf\left\{u > 0 \text{ with } \mathbb{E}_\xi[\Phi(\|z(\xi, \zeta)\|/u)] \leqslant 1\right\}. \tag{5.1}$$

Then $\|Z\|_\Phi \leqslant \big\| \|Z\|_{\Phi,\xi} \big\|_\Phi$.

Proof. One needs to prove that $\mathbb{E}[\Phi(Z/\|\|Z\|_{\Phi,\xi}\|_\Phi)] \leqslant 1$:

$$\begin{aligned}
\mathbb{E}\left[\Phi\left(\frac{Z}{\|\|Z\|_{\Phi,\xi}\|_\Phi}\right)\right] &\leqslant \mathbb{E}\left[\Phi\left(\frac{Z}{\|Z\|_{\Phi,\xi}}\frac{\|Z\|_{\Phi,\xi}}{\|\|Z\|_{\Phi,\xi}\|_\Phi}\right)\right] \\
&\leqslant \mathbb{E}\left[\Phi\left(\frac{Z}{\|Z\|_{\Phi,\xi}}\right)\Phi\left(\frac{\|Z\|_{\Phi,\xi}}{\|\|Z\|_{\Phi,\xi}\|_\Phi}\right)\right].
\end{aligned}$$

The last inequality follows from (2.1). By independence of ξ and ζ and by (5.1)

$$\begin{aligned}
\mathbb{E}\left[\Phi\left(\frac{Z}{\|\|Z\|_{\Phi,\xi}\|_\Phi}\right)\right] &\leqslant \mathbb{E}\left[\Phi\left(\frac{\|Z\|_{\Phi,\xi}}{\|\|Z\|_{\Phi,\xi}\|_\Phi}\right)\mathbb{E}_\xi\left[\Phi\left(\frac{Z}{\|Z\|_{\Phi,\xi}}\right)\right]\right] \\
&\leqslant \mathbb{E}\left[\Phi\left(\frac{\|Z\|_{\Phi,\xi}}{\|\|Z\|_{\Phi,\xi}\|_\Phi}\right)\right].
\end{aligned}$$

We conclude by using the definition of the norm $\|\cdot\|_\Phi$. □

Lemma 5.3. *If the Orlicz function Φ satisfies (2.1) then for any E-valued random variable X we have $\|X\|_1 \leqslant \|X\|_\Phi$.*

Proof. Using Jensen's inequality, we obtain

$$\mathbb{E}\left[\Phi\left(\frac{\|X\|}{\|X\|_1}\right)\right] \geqslant \Phi(1).$$

Note that $\Phi(1) \leqslant \Phi(1)^2$ by (2.1) and then that $\Phi(1) \geqslant 1$. We conclude that $\|X\|_1 \leqslant \|X\|_\Phi$ by using the definition of the norm $\|\cdot\|_\Phi$. □

Lemma 5.4. *Let $u_0 \geqslant 0$ and $(u_n)_{n\in\mathbb{Z}}$ be a real sequence such that $|u_n| \leqslant u_0$ if $n < 0$. Assume that*

$$u_n = \sum_{i=1}^{p} a_i u_{n-i}, \quad \forall n \geqslant 0, \tag{5.2}$$

where $a_1, \ldots, a_p$ are fixed non-negative numbers with $a = \sum_{i=1}^{p} a_i < 1$. Then,

$$\sup_{k\geqslant n} u_k \leqslant a^{n/p} u_0, \quad \forall n \geqslant 0.$$

Proof. By a recursion argument, one first shows that $\sup_{k\leqslant n} u_k \leqslant u_0$. Then $(u_n)_{n\in\mathbb{N}}$ is bounded by u_0. Let $v_n = \sup_{k\geqslant n} u_k$ for $n \in \mathbb{Z}$. Using the relation (5.2), we get $v_n \leqslant a v_{n-p}$ for all $n \geqslant 0$. Then recursively $v_n \leqslant a^{-[-n/p]} v_{n+p[-n/p]}$. From $|u_n| \leqslant u_0$ if $n < 0$, $v_{n+p[-n/p]} = v_0 = u_0$ because $n + p[-n/p] \leqslant 0$. The result follows from $-[-n/p] \geqslant n/p$. □

5.2. *p*-Markov stationary approximations

In order to construct a solution to (1.1) we consider, for each fixed $p \geqslant 0$ and $q > 0$, the p-Markov process $(X_{p,q,t})_{t\geqslant 0}$ defined by $X_{p,q,t} = 0$ for $t \leqslant -q$ and the recurrence equation

$$X_{p,q,t} = F(X_{p,q,t-1}, \ldots, X_{p,q,t-p}, 0, 0, \ldots; \xi_t) \quad \text{if } t > q. \tag{5.3}$$

Using the notation of Lemma 5.1 with $\xi = \xi_0$ and $\zeta = (X_{p,q,-1}, X_{p,q,-2}, \ldots)$ and $z(\xi,\zeta) = F(\zeta,\xi)$, the Lipschitz condition (3.1) implies that

$$\|X_{p,q+1,0} - X_{p,q,0}\|_{\Phi,\xi} \leqslant \sum_{i=1}^{p} a_i \|X_{p,q+1,-i} - X_{p,q,-i}\|.$$

Applying Lemma 5.2,

$$\begin{aligned}\left\|X_{p,q+1,0} - X_{p,q,0}\right\|_\Phi &\leqslant \|\, \|X_{p,q+1,0} - X_{p,q,0}\|_{\Phi,\xi}\, \|_\Phi \\ &\leqslant \left\| \sum_{i=1}^{p} a_i \|X_{p,q+1,-i} - X_{p,q,-i}\| \right\|_\Phi \\ &\leqslant \sum_{i=1}^{p} a_i \left\|X_{p,q+1,-i} - X_{p,q,-i}\right\|_\Phi \\ &\leqslant \sum_{i=1}^{p} a_i \left\|X_{p,q+1-i,0} - X_{p,q-i,0}\right\|_\Phi .\end{aligned}$$

The last inequality follows from the fact that by the definition of $X_{p,q,-i}$ and $X_{p,q-i,0}$, these quantities have the same law for each triplet of positive integers (p, q, i). We now consider $v_n = \left\| X_{p,n+1,0} - X_{p,n,0} \right\|_\Phi$ for $n \in \mathbb{Z}$, with $v_n = 0$ if $n < 0$. For $n > 0$

$$v_n \leqslant \sum_{i=1}^{p} a_i v_{n-i}.$$

From Lemma 5.4 we obtain

$$v_n \leqslant a^{n/p} v_0 \leqslant a^{n/p} \|X_{p,1,0}\|_\Phi \leqslant a^{n/p} \|F(0, 0, \ldots; \xi_t)\|_\Phi \leqslant a^{n/p} \mu_\Phi.$$

Hence, for each p, $(X_{p,n,0})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{L}^\Phi$; it converges to some $X_{p,0} \in \mathbb{L}^\Phi$. From its construction, it is clear that $X_{p,n,0}$ is measurable with respect to the σ-algebra generated by $\{\xi_t, t \leqslant 0\}$. The $\mathbb{L}^\Phi$-convergence ensures that this is also the case for $X_{p,0}$. Hence there exists some measurable function H_p such that $X_{p,0} = H_p(\xi_0, \xi_{-1}, \ldots)$. As $n \uparrow \infty$, a continuity argument on F implies that $X_{p,0} = F(X_{p,-1}, \ldots, X_{p,-p}, 0, 0, \ldots; \xi_0)$ and shifting the lag $t \in \mathbb{Z}$ leads to the equalities

$$X_{p,t} = H_p(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots) = F(X_{p,t-1}, \ldots, X_{p,t-p}, 0, 0, \ldots; \xi_t).$$

Then the sequence $(X_{p,t})_{t\in\mathbb{Z}}$ is a stationary solution of the recurrence equation (5.3) for each $p \geqslant 0$.

Consider

$$\mu_{\Phi,p} = \|X_{p,t}\|_\Phi, \qquad \Delta_{\Phi,p,t} = \|X_{p+1,t} - X_{p,t}\|_\Phi.$$

The definition of $\mu_{\Phi,p}$ given here for $p > 0$ extends to $p = 0$ since $X_{0,t} = F(0, 0, \ldots; \xi_t)$ satisfies $\|X_{0,t}\|_\Phi = \mu_\Phi$ by Eq. (3.3).

Lemma 5.5. *Assume conditions (3.2) and (3.3) hold for some Orlicz function Φ satisfying (2.1). Then*

$$\mu_{\Phi,\infty} = \sup_{p\geqslant 0} \mu_{\Phi,p} \leqslant \frac{\mu_\Phi}{1-a} \quad \text{and} \quad \Delta_{\Phi,p} = \sup_{t\in\mathbb{Z}} \Delta_{\Phi,p,t} \leqslant a_{p+1} \frac{\mu_\Phi}{(1-a)^2}.$$

Proof. From Eq. (3.2), we have that

$$\mu_{\Phi,p} \leqslant \|X_{p,t} - X_{0,t}\|_\Phi + \mu_\Phi \leqslant \sum_{j=1}^{p} a_j \|X_{p,t-j}\|_\Phi + \mu_\Phi \leqslant \mu_{\Phi,p} \sum_{j=1}^{p} a_j + \mu_\Phi,$$

and hence $\mu_{\Phi,p} \leqslant (1-a)^{-1}\mu_\Phi$ and $\mu_{\Phi,\infty} \leqslant (1-a)^{-1}\mu_\Phi$ follow. In a similar way, we obtain the inequalities

$$\begin{aligned}
\Delta_{\Phi,p,t} &= \big\| F(X_{p+1,t-1}, \ldots, X_{p+1,t-p-1}, 0, 0, \ldots; \xi_t) \\
&\quad - F(X_{p,t-1}, \ldots, X_{p,t-p}, 0, 0, \ldots; \xi_t) \big\|_\Phi \\
&\leqslant \sum_{j=1}^{p} a_j \|X_{p+1,t-j} - X_{p,t-j}\|_\Phi + a_{p+1} \|X_{p+1,t-p-1}\|_\Phi \\
&\leqslant \sum_{j=1}^{p} a_j \Delta_{\Phi,p,t-j} + a_{p+1} \|X_{p+1,0}\|_\Phi.
\end{aligned}$$

This implies that $\Delta_{\Phi,p} \leqslant a_{p+1}(1-a)^{-1}\mu_{\Phi,p+1}$ and the result of Lemma 5.5 is shown. $\square$

5.3. Proof of the existence of a solution to (1.1)

Note first that Lemma 5.5 implies that $X_{p,t} \to_{p\to\infty} X_t$ in $\mathbb{L}^{\Phi}$ since this space is complete. The continuity of F ensures that X_t is a solution of Eq. (1.1). Furthermore, as a limit in $\mathbb{L}^{\Phi}$ of strictly stationary processes, X_t is also stationary (in law) and $\|X_t\|_{\Phi} < \infty$. Finally, $X_t = H(\xi_t, \xi_{t-1}, \ldots)$ is the limit in $\mathbb{L}^{\Phi}$ of $X_{p,t} = H_p(\xi_t, \xi_{t-1}, \ldots)$.

5.4. Proof of the weak dependence properties

The weak dependence property of a solution to (1.1) is formulated in terms of the $\mathbb{L}^1$-norm in the definition of the coefficients τ. As shown in Lemma 5.3, $\|X\|_1 \leqslant \|X\|_{\Phi}$ for any E-valued random variable X. Then assumptions (3.2) and (3.3) are always satisfied, replacing $\|\cdot\|_{\Phi}$ with $\|\cdot\|_1$. We first prove Corollary 3.1:

Proof. We use coupling techniques to evaluate the coefficients τ; see Section 2. Let $(\xi'_t)_{t\in\mathbb{Z}}$ be an independent copy of $(\xi_t)_{t\in\mathbb{Z}}$. We define the process $(X^*_{p,t})_{t\in\mathbb{Z}}$ as

$$X^*_{p,t} = \begin{cases} F(X^*_{p,t-1}, \ldots, X^*_{p,t-p}, 0, 0, \ldots; \xi'_t), & \text{for } t \leqslant 0; \\ F(X^*_{p,t-1}, \ldots, X^*_{p,t-p}, 0, 0, \ldots; \xi_t), & \text{for } t > 0; \end{cases}$$

Using arguments similar to those of Section 5.2, there exists a sequence of measurable variables with respect to the σ-algebra generated by $\xi'_t, t \leqslant 0$ denoted by $(X^*_{p,n,0})_{n\in\mathbb{N}}$ such that it converges in $\mathbb{L}^{\Phi}$ to $X^*_{p,0} \in \mathbb{L}^{\Phi}$. The $\mathbb{L}^{\Phi}$-convergence ensures that $X^*_{p,0}$ are also measurable variables with respect to the σ-algebra generated by $\xi'_t, t \leqslant 0$. Then, by definition of $\xi'_t, t \leqslant 0$, $X^*_{p,0}$ is independent of $X_{p,0}$. If there exists a non-increasing function $\delta_p(r)$ of r such that $\|X_{p,r} - X^*_{p,r}\|_1 \leqslant \delta_p(r)$, we have $\tau_{p,r} \leqslant \delta_p(r)$. This follows from the coupling property of weak dependence coefficients τ explained in [8].

Assumption (3.2) and Lemma 5.3 yield

$$\|X_{p,r} - X^*_{p,r}\|_1 \leqslant \sum_{i=1}^{p} a_i \|X_{p,r-i} - X^*_{p,r-i}\|_1.$$

Defining $w_r = \|X_{p,r} - X^*_{p,r}\|_1$ for $r \in \mathbb{Z}$, we again use Lemma 5.4 and the relation $\|F(0,0,\ldots;\xi_0)\|_1 = \mu_1$ to obtain

$$w_r \leqslant a^{r/p} w_0 \leqslant 2\mu_1 a^{r/p} \leqslant 2\frac{\mu_1}{1-a} a^{r/p}.$$

Now choosing $\delta_p(r) := 2\mu_1(1-a)^{-1}a^{r/p}$ leads to the result of Corollary 3.1. $\square$

Now we finish the proof of Theorem 3.1, defining the process $(X^*_t)_{t\in\mathbb{Z}}$ as the solution of the equations

$$X^*_{t\in\mathbb{Z}} = \begin{cases} F(X^*_{t-1}, X^*_{,t-2}, \ldots; \xi'_t), & \text{for } t \leqslant 0; \\ F(X^*_{t-1}, X^*_{t-2}, \ldots; \xi_t), & \text{for } t > 0; \end{cases}$$

We remark that $(X^*_t)_t$ is also a stationary chain with infinite memory. Lemma 5.5 gives

$$\|X_r - X_{p,r}\|_1 \leqslant \sum_{k=p}^{\infty} \Delta_{1,k} \leqslant \frac{\mu_1}{(1-a)^2} \sum_{k=p}^{\infty} a_{k+1}.$$

The same bound holds for the quantity $\|X_r^* - X_{p,r}^*\|_1$. For each integer p,

$$\|X_r - X_r^*\|_1 \leqslant \|X_r - X_{p,r}\|_1 + \|X_{p,r} - X_{p,r}^*\|_1 + \|X_r^* - X_{p,r}^*\|_1$$
$$\leqslant 2\frac{\mu_1}{1-a}\left(a^{r/p} + \sum_{k=p+1}^{\infty}\frac{a_k}{1-a}\right).$$

Because this bound is non-increasing with r, we conclude the weak dependence properties in Theorem 3.1 by using the coupling technique.

5.5. *Proof of Theorem 3.2*

First we recall the assumption (D(q)) of [6] for $q > 1$,

$$\int_0^{\|X_0\|_1} ((\tau/2)^{-1}(u))^{q-1} Q^{q-1} \circ G(u)\mathrm{d}u < \infty, \tag{D(q)}$$

where $(\tau/2)^{-1}(u) = \inf\{k \in \mathbb{N}/\ \tau(k) \leqslant 2u\}$. Here Q denotes the generalized inverse of the tail function $x \mapsto \mathbb{P}(|X_0| > x)$ and G the inverse of $x \mapsto \int_0^x Q(u)\mathrm{d}u$. Dedecker and Doukhan proved in [6] the SLLN and the CLT under (D(q)) for respectively $1 < q < 2$ and $q = 2$. The SIP is proved in [8] under (D(q)) for $q = 2$. Write $A(p) = \sum_{j\geqslant p} a_j$ and A^{-1} its generalized inverse $A^{-1}(u) = \inf\{k \in \mathbb{N}/\ A(u) \leqslant u\}$,

$$\Psi_q(x) = \Phi\left(x^{1/(q-1)}\right)/x^{1/(q-1)} \quad \text{and} \quad \Psi_q^*(x) = \sup_{y\geqslant 0}\{xy - \Psi_q(y)\}.$$

Noticing that $A^{-1}(u) = k$ on $]A(k-1); A(k)]$ and that $\widetilde{\Phi}_q(x) = \Psi_q^*(x^{q-1})$, there exists $C > 0$ such that

$$\int_0^a \widetilde{\Phi}_q\left(c_0(A^{-1}(u)-1)\ln(u)\right)\mathrm{d}u \leqslant C\sum_{k\geqslant 1} a_k \widetilde{\Phi}_q\left(c_0 k\left(1 - \mathbb{1}_{\left\{\sum_{j\geqslant k} a_j > 0\right\}}\ln\left(\sum_{j\geqslant k} a_j\right)\right)\right).$$

Then assumption (3.4) implies that we work under the condition

$$\int_0^a \Psi_q^*\left(\left(c_0(A^{-1}(u)-1)\ln(u)\right)^{q-1}\right)\mathrm{d}u < \infty. \tag{5.4}$$

We want to prove that condition (5.4) implies (D(q)) for all $q > 1$. The first step is to prove the bound

$$(\tau/2)^{-1}(u) \leqslant \left[\left(A^{-1}\left(\frac{1-a}{2\mu_1}u\right) - 1\right)\frac{\ln\left(\frac{1-a}{2\mu_1}u\right)}{\ln a}\right], \tag{5.5}$$

Theorem 3.1 gives $(\tau/2)^{-1}(u) \leqslant \inf B$ with

$$B = \left\{k \in \mathbb{N} \text{ such that } \exists p \geqslant 1 \text{ with } \frac{\mu_1}{1-a}\left(a^{k/p} + A(p+1)\right) \leqslant u\right\}.$$

Set $v = (1-a)(2\mu_1)^{-1}u$; the integer $p^* = A^{-1}(v) - 1$ is close to the infimum of B. Then all integers k with $a^{k/p^*} \leqslant v$ belong to B, for instance $k^* = [(A^{-1}(v) - 1)\ln v/\ln a]$ which is then larger than $(\tau/2)^{-1}(u)$ by definition. Observe that $A^{-1}(v) = 1$ as soon as $v \geqslant a$; thus $[(A^{-1}(v) - 1)\ln v/\ln a] = 0$ for $v \geqslant a$.

Using this estimate of $(\tau/2)^{-1}$ in (5.5), condition (D(q)) holds if

$$\int_0^a \left[\frac{(A^{-1}(v) - 1)\ln v}{\ln a}\right]^{p-1} Q^{p-1} \circ G\left(\frac{2\mu_1}{1-a}v\right) \mathrm{d}v < \infty. \tag{5.6}$$

Let $\widetilde{\Psi}$ be an Orlicz function and $\widetilde{\Psi}^*(x) = \sup_{y>0}\{xy - \widetilde{\Psi}(y)\}$ be its Young dual function. For any functions f and g, Young's inequality gives

$$\int_0^a f(x)g(x)\mathrm{d}x \leqslant 2\inf\left\{c > 0 \text{ with } \int_0^a \widetilde{\Psi}\left(\frac{f(x)}{c}\right)\mathrm{d}x \leqslant 1\right\}$$
$$\times \inf\left\{c > 0 \text{ with } \int_0^a \widetilde{\Psi}^*\left(\frac{g(x)}{c}\right)\mathrm{d}x \leqslant 1\right\}.$$

In the following we apply this inequality with $\widetilde{\Psi} = K\Psi_p$ for some $K > 0$, $f(x) = Q^{p-1} \circ G(2\mu_1(1-a)^{-1}x)$ and $g(x) \leqslant ((A^{-1}(x) - 1)\ln(1/x)(-\ln a)^{-1})^{q-1}$. Note that the Young dual function is here $\widetilde{\Psi}^*(x) = K\Psi_q^*(x/K)$ and then $\int_0^a f(x)g(x)\mathrm{d}x$ is equal to the left hand side term (5.6) up to the choice of the constant $K > 0$; see below. In view of Young's inequality, the first term in the bound of (5.6) thus expresses as the infimum over $c > 0$ such that

$$K\frac{1-a}{2\mu_1}\int_0^{\|X_0\|_1} \frac{\Phi\left(Q \circ G\left(u\right)/c\right)}{Q \circ G\left(u\right)/c}\mathrm{d}u \leqslant 1.$$

Replacing $G(u)$ with x, one obtains the simpler inequality

$$K\frac{1-a}{2\mu_1}\int_0^1 \Phi\left(\frac{Q(x)}{c}\right)c\mathrm{d}x = K\frac{1-a}{2\mu_1}c\mathbb{E}\Phi\left(\frac{|X_0|}{c}\right) \leqslant 1.$$

The last equality is set using the definition of $Q(x)$. If assumption (3.3) holds, the last inequality is satisfied for $K = 2\mu_1\mu_\Phi^{-1}$ and $c = \mu_\Phi(1-a)^{-1}$.

The second term of the Young inequality is expressed as the infimum over $c > 0$ such that

$$K\int_0^a \Psi_q^*\left(\frac{((A^{-1}(x) - 1)\ln(1/x))^{q-1}}{K(-\ln a)^{q-1}c}\right)\mathrm{d}x \leqslant 1. \tag{5.7}$$

Because $\widetilde{\Phi}_q(x) = \Psi_q^*(x^{q-1})$ we check that

$$0 < \frac{\int_0^a \widetilde{\Phi}_q\left(c_0(A^{-1}(u) - 1)\ln(u)\right)\mathrm{d}u \vee 1}{(K \wedge 1)(-\ln a)^{p-1}} =: c_1$$

satisfies the relation (5.7). It is obvious by (5.4) that $c_1 < \infty$ and then we have proved the implications

(3.4) with $q > 1 \Rightarrow$ (5.4) with $q > 1 \Rightarrow$ (D(q)).

This ends the proof as the results of Theorem 3.2 are versions of the results in [6,8] that hold under assumption (D(q)).

5.6. Proof of Proposition 3.1

Let n be a fixed integer and $s_n \leqslant n-1$. Let $(X_t)_{t\in\mathbb{Z}}$ be the stationary solution of $X_t = F(X_{t-1}, X_{t-2}, 0, 0, \ldots; \xi_t)$. The Lipschitz assumption (3.1) implies for $1 \leqslant k \leqslant n$

$$\left\| \widetilde{X}_k - X_k \right\|_\Phi \leqslant \sum_{i=1}^{k-1} a_i \left\| \widetilde{X}_{k-i} - X_{k-i} \right\|_\Phi + \sum_{i\geqslant k} a_i \| X_0 - c_i \|_\Phi .$$

The sequence $v_k = \left\| \widetilde{X}_{k+1} - X_{k+1} \right\|_\Phi$, $k = 1, 2, \ldots$, satisfies the recursion

$$v_k \leqslant \sum_{j=1}^{k} a_j v_{k-j} + u_k \quad \text{for all } k \geqslant 1,$$

where $u_k = (\|X_0\|_\Phi + \overline{c}) \sum_{j>k} a_j$ for $k \geqslant 1$. Notice that $u_k \downarrow_{k\to\infty} 0$. We first prove the boundedness of $(v_k)_{k\in\mathbb{N}}$. Let ℓ be a fixed integer. For all k such that $\ell \geqslant k$, $v_k \leqslant a \sup_{i\leqslant \ell} v_i + u_1$. We deduce that $\sup_{i\leqslant\ell} v_i \leqslant u_1$. Finally $\|v\|_\infty \leqslant a\|X_0\|_\Phi/(1-a)$.
Now for all integers $k, s \geqslant 1$ such that $\ell \geqslant k + s$,

$$v_\ell \leqslant \sum_{j=1}^{k} a_j v_{\ell-j} + \sum_{j=k+1}^{\ell} a_j v_{\ell-j} + u_\ell \leqslant a \sup_{j\geqslant s} v_j + \|v\|_\infty \sum_{j=k+1}^{\infty} a_j + u_{k+s}.$$

This inequality holds for all $\ell \geqslant k + s$. Then

$$\sup_{j\geqslant k+s} v_j \leqslant a \sup_{j\geqslant s} v_j + \|v\|_\infty \sum_{j=k+1}^{\infty} a_j + u_k .$$

We deduce that

$$\sup_{j\geqslant nk} v_j \leqslant a^n \|v\|_\infty + \frac{1}{1-a} \left(\|v\|_\infty \sum_{j=k+1}^{\infty} a_j + u_k \right).$$

Using the inequality $\|v\|_\infty \leqslant a\|X_0\|_\Phi/(1-a)$, one gets the result.

Acknowledgements

We are deeply grateful to Alain Latour, and Thomas Mikosch who made a critical review of the drafts and with whom we have worked on the final version of this paper. We also wish to thank Jérôme Dedecker for important comments relating τ-weak dependence conditions as well as Lionel Truquet who pointed out the example of Galton–Watson processes with immigration. We finally wish to thank an anonymous referee whose remarks considerably helped us to improve our results.

Appendix

We give below general conditions for the existence and the boundedness of joint densities of affine models defined in Section 4. Thus we extend the results for bilinear models given in [14].

Proposition 5.1 (*Regularity of Affine Models*). *Here $E = E' = \mathbb{R}^d$ for some $d \geqslant 1$. Suppose that the innovations $(\xi_t)_{t\in\mathbb{Z}}$ in the model (4.2) have a common bounded marginal* *density* f_ξ. *Moreover, if* $\inf_{(x_j)_{j>0}} \det M((x_j)_{j>0}) = \underline{M} > 0$, *the marginal densities* $f_{X_1,\dots,X_n}$ *of* $(X_1,\dots,X_n)$ *exist for all* $n > 0$ *and satisfy*

$$\|f_{X_1,\dots,X_n}\|_\infty \leqslant \underline{M}^{-n}\|f_\xi\|_\infty^n.$$

Proof. The solution $X_t = H(\xi_t, \xi_{t-1}, \dots)$ obtained in Section 5.3 is independent of $(\xi_j)_{j>t}$. If G_1 is a bounded continuous function on E with values in $\mathbb{R}$, it holds that

$$\begin{aligned}\mathbb{E}G(X_1) &= \mathbb{E}G_1(M(X_0,\dots)\xi_1 + f(X_0, X_{-1},\dots))\\ &= \int\int G_1(M(u)s_1 + f(u)) f_\xi(s_1)\mathrm{d}s_1 \mathbb{P}_{(X_0,X_{-1},\dots)}(\mathrm{d}u)\\ &\leqslant \underline{M}\int\int G(x_1) f_\xi(M^{-1}(u)(x_1 - f(u)))\mathbb{P}_{(X_0,X_{-1},\dots)}(\mathrm{d}u)\mathrm{d}s_1.\end{aligned}$$

The last inequality follows by the substitution $M(u)s_1 + f(u) = x_1$ valid under the assumption $\inf_{(x_j)_{j>0}} \det M((x_j)_{j>0}) = \underline{M} > 0$ ensuring that $M(u)$ is invertible for all u. We obtain

$$f_{X_1}(x_1) \leqslant \underline{M}^{-1}\int f_\xi(M^{-1}(u)(x_1 - f(u)))\mathbb{P}_{(X_0,X_{-1},\dots)}(\mathrm{d}u) \leqslant \underline{M}^{-1}\|f_\xi\|_\infty.$$

We proceed by induction for the cases $n \geqslant 2$. Assume that $\|f_{X_1,\dots,X_{n-1}}\|_\infty \leqslant \underline{M}^{-(n-1)}\|f_\xi\|_\infty^{n-1}$ is satisfied. Let G_n be a bounded continuous function on E^n with value in $\mathbb{R}$; one has

$$\begin{aligned}&\mathbb{E}G_n(X_1,\dots,X_n) = \mathbb{E}G_n(X_1,\dots,X_{n-1}, M(X_{n-1},X_{n-2},\dots)\xi_n + f(X_{n-1},X_{n-2},\dots))\\ &= \int\int\int G_n(x_1,\dots,x_{n-1}, M(x_{n-1},\dots,x_1,u)s_n + f(x_{n-1},\dots,x_1,u))\\ &\quad f_\xi(s_n)\mathrm{d}s_n f_{(X_1,\dots,X_{n-1})}(x_1,\dots,x_{n-1})\mathrm{d}x_1\cdots\mathrm{d}x_{n-1}\mathrm{d}P_{(X_0,X_{-1},\dots|X_1,\dots,X_{n-1})}(u).\end{aligned}$$

The substitution $M(x_{n-1},\dots,x_1,u)s_n + f(x_{n-1},\dots,x_1,u) = x_n$ yields

$$\begin{aligned}f_{X_1,\dots,X_n}(x_1,\dots,x_n) &\leqslant \underline{M}^{-1}\int\int f_\xi(M^{-1}(x_{n-1},\dots,x_1,u)(x_n - f(x_{n-1},\dots,x_1,u)))\\ &\quad\times f_{(X_1,\dots,X_{n-1})}(x_1,\dots,x_{n-1})\mathrm{d}x_1\cdots\mathrm{d}x_{n-1}\mathrm{d}P_{(X_0,X_{-1},\dots|X_1,\dots,X_{n-1})}(u).\end{aligned}$$

Together with the induction assumption $\|f_{X_1,\dots,X_{n-1}}\|_\infty \leqslant \underline{M}^{-(n-1)}\|f_\xi\|_\infty^{n-1}$, this last inequality yields $\|f_{X_1,\dots,X_n}\|_\infty \leqslant \underline{M}^{-n}\|f_\xi\|_\infty^n$. □

References

[1] D.W.K. Andrews, Nonstrong mixing autoregressive processes, J. Appl. Probab. 21 (4) (1984) 930–934.
[2] H. Berbee, Chains with infinite connections: Uniqueness and Markov representation, Probab. Theory Related Fields 76 (2) (1987) 243–253.
[3] P. Bougerol, Kalman filtering with random coefficients and contractions, Probab. Theory Related Fields 31 (1993) 942–959.
[4] P. Bühlmann, A.J. Wyner, Variable length Markov chains, Ann. Statist. 27 (2) (1999) 480–513.
[5] F. Comets, R. Fernández, P.A. Ferrari, Processes with long memory: Regenerative construction and perfect simulation, Ann. Appl. Probab. 12 (3) (2002) 921–943.
[6] J. Dedecker, P. Doukhan, A new covariance inequality and applications, Stochastic Process. Appl. 106 (1) (2003) 63–80.
[7] J. Dedecker, P. Doukhan, G. Lang, J.R. León, S. Louhichi, C. Prieur, Weak Dependence, Examples and Applications, in: Lecture Notes in Statistics, vol 190, Springer-Verlag, Berlin, 2007.
[8] J. Dedecker, C. Prieur, Coupling for τ-dependent sequences and applications, J. Theoret. Probab. 17 (4) (2004) 861–855.

[9] P. Diaconis, D. Freedman, Iterated random functions, SIAM Rev. 41 (1) (1999) 45–76.
[10] R.L. Dobrushin, Prescribing a system of random variables by conditional distributions, Theory Probab. Appl. 15 (1970) 458–486.
[11] R.L. Dobrushin, S. Kusuoka, Statistical Mechanics and Fractals, in: Lecture Notes in Mathematics, vol. 1567, Springer-Verlag, New York, 1993.
[12] P. Doukhan, Mixing, in: Lecture Notes in Statistics, vol. 85, Springer-Verlag, New York, 1994.
[13] P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, Stochastic Process. Appl. 84 (2) (1999) 313–342.
[14] P. Doukhan, H. Madre, M. Rosenbaum, ARCH type bilinear weakly dependent models, Statistics 41 (1) (2007) 31–45.
[15] P. Doukhan, P. Massart, E. Rio, The functional central limit theorem for strongly mixing processes, Ann. Inst. H. Poincaré Probab. Statist. 30 (2) (1994) 63–82.
[16] P. Doukhan, G. Teyssière, P. Winant, A LARCH(∞) vector valued process, in: Patrice Bertail, Paul Doukhan, Philippe Soulier (Eds.), Dependence in Probability and Statistics, in: Lectures Notes in Statistics, vol. 187, Springer, New York, 2006.
[17] M. Duflo, Random Iterative Models, in: Applications of Mathematics, vol. 34, Springer-Verlag, Berlin, 1997.
[18] R. Fernández, G. Maillard, Chains with complete connections and one-dimensional Gibbs measures, Electron. J. Probab. 9 (6) (2004) 145–176.
[19] L. Giraitis, R. Leipus, P.M. Robinson, D. Surgailis, LARCH, leverage and long memory, J. Financial Econometrics 2 (2) (2004) 177–210.
[20] M. Iosifescu, S. Grigorescu, Dependence with Complete Connections and its Applications, in: Cambridge Tracts in Mathematics, vol. 96, Cambridge University Press, Cambridge, 1990.
[21] M. Iosifescu, R. Theodorescu, Random Processes and Learning, Springer-Verlag, New York, 1969.
[22] M. Kac, Probability and Related Topics in Physical Sciences, in: Lectures in Applied Mathematics Series, vol. 1a, American Mathematical Society, London, 1959.
[23] O. Kallenberg, Foundations of Modern Probability, in: Probability and its Applications, Springer-Verlag, New York, 1997.
[24] M.A. Krasnoselskii, Y.B. Rutickii, Convex Functions and Orlicz Spaces, Noordhoff Ltd., Groningen, 1961.
[25] L.D. Landau, E.M. Lifshitz, Statistical Physics, 3rd edition, in: Course of Theoretical Physics, vol. 5, Butterworth-Heinemann, Oxford, 1980.
[26] A. Latour, The multivariate GINAR(p) process, Adv. Appl. Prob. 29 (1997) 228–247.
[27] F. Merlevde, M. Peligrad, On the weak invariance principle for stationary sequences under projective criteria, J. Theoret. Probab. 19 (2006) 647–689.
[28] M. Peligrad, S. Utev, A new maximal inequality and invariance principle for stationary sequences, Ann. Probab. 33 (2005) 798–815.
[29] N. Ragache, O. Wintenberger, Convergence rates for density estimators of weakly dependent time series, in: P. Bertail, P. Doukhan, P. Soulier (Eds.), Dependence in Probability and Statistics, in: Lectures Notes in Statistics, vol. 187, Springer, New York, 2006.
[30] E. Rio, Théorie asymptotique des processus aléatoires faiblement dépendants, in: Mathématiques & Applications, vol. 31, Springer-Verlag, Berlin, 2000.
[31] W.B. Wu, Nonlinear system theory: Another look at dependence, Proc. Natl. Acad. Sci. USA 102 (2005) 14150–14154.