submitted

WEAKLY DEPENDENT CHAINS WITH INFINITE MEMORY

By Paul Doukhan, Olivier Wintenberger

LS-CREST and University Paris 1 Panthéon-Sorbonne

The main objective of the paper is to define $\mathbb{L}^m$ $(m \geq 1)$ strictly stationary solutions of infinite memory recurrence equations such as $X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \ldots; \xi_t)$, where $(\xi_t)_{t\in\mathbb{Z}}$ denotes an independent and identically distributed sequence. To this end, we use an appropriate Lipschitz condition which also entails weak dependence properties defined in Dedecker and Prieur (2004). Such models are proved to provide continuous state space and nonlinear extensions of various examples of well known classes of times series.

1. Introduction. Times series analysis is a main research field for application sake. The statistical analysis of times series deeply relies on the underlying model, *e.g.* selection of models is now very popular in statistics. The choice of a model balances several antagonist criteria. We focus here on the principle of parsimony which consists on reducing as much as possible the set of parameters in a model. We apply this principle bearing in mind that the models must not be too restrictive.

Regular stationary second order times series are traditionally represented as infinite autoregressive processes

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + \xi_t, \qquad t \in \mathbb{Z} \tag{1.1}$$

with $(\xi_t)_{t\in\mathbb{Z}}$ a second order stationary white noise. This representation has motivated the introduction of $ARMA$ processes. In Econometrics, Industry *e.g.* those models are widely used.

The expression of a process from its own history motivates a lot of forecast attempts. Assuming linearity in the representation (1.1) induces that innovations are second-order stationary but not strictly stationary. On the contrary, non-linearity allow the use of strictly stationary inputs. Markov models first introduce non-linearity. Various applications (*e.g.* Finance, Hydrodynamics, Physics, Electromagnetism, see Landau and Lifschitz (1969), or Dobrushin and Kusuoka (1993)) use such representations.

AMS 2000 subject classifications: Primary 60K99; Secondary 70G70, 60J05

Keywords and phrases: stationary solutions, weak dependence, coupling

Dobrushin (1970)'s conditions are widely used to define random fields from their conditional distributions; we defer the reader to Georgii (1988) for details. Such conditions need regularity of the marginal distributions. They are closely related with sharp mixing conditions (see Doukhan, 1994). To allow weaker dependence behavior, the Markov property is relaxing. Berbee (1987) obtained an existence condition for discrete state spaces in such non Markovian context. Comets *et al.* (2002) and Fernandez and Maillard (2006) improved this condition. Iosifescu and Teodorescu (1969) introduced the underlying stationary model. General random systems with complete connections (RSCC) are defined through their marginal conditional distributions in an extended Markovian way. Bühlmann and Wyner (1999) introduce Variable Length Markov Chains (VLMC) as an alternative model: triangular arrays $X_{1,n}, \ldots, X_{n,n} \in \mathcal{P}_n$, are considered for some set $\mathcal{P}_n$ set of p_n-Markov chains (here ergodicity replaces the condition of stationarity). Omitting both linearity and Markov assumptions for the representation of a random phenomenon, those models are widely used in the fields of particle systems or in DNA data analysis.

We place us in the different context of continuous state spaces. Models both non Markovian and non linear already exist. In the following, $(\xi_t)_{t\in\mathbb{Z}}$ is called the innovations or inputs of the system. It is a sequence of independent and identically distributed (iid) random variables with values in a probability space E. Attractive examples of weakly dependent processes are Bernoulli shifts defined as (cf. [13], and also [26] for references)

$$X_t = H(\xi_{t-j}, j \in \mathbb{N}), \text{ for } t \in \mathbb{Z} \text{ and } H : E^{\mathbb{N}} \to \mathbb{R}^d \text{ measurable.}$$

Such Bernoulli shifts look general (see [7]) and this was even conjectured by Wiener that any stationary sequence writes this way. However, those models do not fit any idea of parsimony, *e.g.* Volterra chaotic models

$$H(x_0, x_1 \ldots) = a_0 + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k=0}^{\infty} a_{k;j_1,\ldots,j_k} x_{j_1} \cdots x_{j_k}$$

involve a very large family of parameters $(a_{k;j_1,\ldots,j_k})_{k,j_1,\ldots,j_k\geq 0}$. Moreover, even if the functional H exists, it remains often not explicit.
We introduce a new and general class of models corresponding to the dynamic behavior of most of the known times series which extends on the Markov property. *Chains with infinite memory* are stationary solutions of equations

$$X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \ldots; \xi_t), \; F : (\mathbb{R}^d)^{\mathbb{N}} \times E \to \mathbb{R}^d \text{ measurable.} \tag{1.2}$$

Such models may be seen as an extension of bilinear models in [14] or $LARCH(\infty)$ models in [15]. Examples in § 3 prove that parsimony is now present for such representations, reducing considerably the family of model parameters.

After some notations we shall motivate more accurately the introduction of such models. Existence, uniqueness and weak dependence properties are derived in our main theorem. Our proof is given in two steps. We first describe the properties of a Markov approximating system in order to prove our main result in two distinct sections.

2. Main results.

2.1. *Definitions.* We need to introduce some notations in order to define the weak dependence coefficients used below.

- The sequence $(\xi_t)_{t\in\mathbb{Z}}$ is iid and takes values in a probability space E.
- In the paper we shall consider a norm $\|\cdot\|$ on $\mathbb{R}^d$ ($d \geq 1$ is a fixed integer). If any confusion is possible, it will be more precisely written with a subscript, *e.g.* the norm on $\mathbb{R}^d$ may also be written $\|\cdot\|_{\mathbb{R}^d}$.
- For a random variable $Z \in \mathbb{R}^d$ we denote $\|Z\|_m = \left(\mathbb{E}\|Z\|^m_{\mathbb{R}^d}\right)^{1/m}$.
- For $h : \mathbb{R}^d \to \mathbb{R}$ we set $\|h\|_\infty = \sup_{x\in\mathbb{R}^d} |h(x)|$ and

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

- $\Lambda_1(\mathbb{R}^d)$ is the set of functions $h : \mathbb{R}^d \to \mathbb{R}$ such that $\mathrm{Lip}\,(h) \leq 1$.

We now recall the notion of weak dependence used in our model.

Definition 2.1 (Dedecker, Prieur, 2004). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-algebra and X be a random variable with values in $\mathbb{R}^d$. Assume that $\mathbb{E}\|X\| < +\infty$, we define the coefficient τ*

(2.1)
$$\tau(\mathcal{M}, X) = \left\|\sup\left\{\left|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\right|,\ f \in \Lambda_1(\mathbb{R}^d)\right\}\right\|_1$$

An easy way to calculate this coefficient is based on a coupling argument (see [8] for details); assume that the probability space $(\Omega, \mathcal{A})$ is rich enough, then there exists X^* distributed as X, independent of $\mathcal{M}$, and such that $\tau(\mathcal{M}, X) = \mathbb{E}\|X - X^*\|$; the relation $\tau(\mathcal{M}, X) \leq \mathbb{E}\|X - X^*\|$ always holds.

Consider an $\mathbb{R}^d$ valued stationary time series $(X_t)_{t\in\mathbb{Z}}$. The definition of τ allows to evaluate the dependence between the past of the sequence $(X_t)_{t\in\mathbb{Z}}$

and its k-tuples in the future. More precisely, consider the norm $\|x-y\| = \|x_1-y_1\|+\cdots+\|x_k-y_k\|$ on $\left(\mathbb{R}^d\right)^k$, we set $\mathcal{M}_p=\sigma(X_t, t\le p)$ and

$$\begin{aligned}\tau_k(r) &= \max_{1\le l\le k}\frac{1}{l}\sup\{\tau(\mathcal{M}_p,(X_{j_1},\ldots,X_{j_l})), p+r\le j_1<\cdots<j_l\},\\ \tau_\infty(r) &= \sup_{k>0}\tau_k(r).\end{aligned}$$

For simplicity sake we will denote below $\tau_\infty(r)=\tau_r$ avoiding any confusion. Firstly, remark that those coefficients do not depend on the value of p thanks to stationarity. Secondly, we quote that those coefficients are causal ones because of the non symmetric role played by the past and the future of the process. We say that a time series $(X_t)_{t\in\mathbb{Z}}$ is τ-weakly dependent when the sequence of its coefficients τ_r tends to 0 as r tends to infinity. We refer to Dedecker and Prieur (2005) for asymptotic results in this setting.

2.2. *The model.* We recall the weak dependence properties of Bernoulli shifts

PROPOSITION 2.1 (Dedecker, Prieur, 2004). *Let $(\xi'_j)_{j\in\mathbb{Z}}$ be distributed as $(\xi_j)_{j\in\mathbb{Z}}$ and independent of it. Assume that $H: E^{\mathbb{N}^*}\to\mathbb{R}^d$ is mesurable and that there exists a sequence $\delta_r\downarrow 0$ as $r\to\infty$ such that*

$$\mathbb{E}\left\|H\left(\xi_{r-j}, j\in\mathbb{N}^*\right)-H\left(\xi_{r-1},\ldots,\xi_1,\xi'_0,\xi'_{-1},\ldots\right)\right\|\le\delta_r, \tag{2.2}$$

then $(X_t)_{t\in\mathbb{Z}}=H(\xi_{t-j}, j\in\mathbb{N})$ is τ-weakly dependent with $\tau_r\le\delta_r$.

The function H is usually unknown and we claim that the equation (1.2) is physically more significant.

Below $F:(\mathbb{R}^d)^{\mathbb{N}}\times E\to\mathbb{R}^d$ denotes a measurable function and $m\ge 1$ is a real number. In order to prove existence of a solution of the equation (1.2), we assume that there exists a sequence $\{a_j\}_{j\in\mathbb{N}^*}$ with $\sum_{j=1}^\infty a_j=a<1$ such that

$$\mu_0=\|F(0,0,0,\ldots;\xi_0)\|_m < \infty, \tag{2.3}$$

$$\|F(x_1,x_2,x_3,\ldots;\xi_0)-F(y_1,y_2,y_3,\ldots;\xi_0)\|_m \le \sum_{j=1}^\infty a_j\|x_j-y_j\|, \tag{2.4}$$

for any $(x_1,x_2,x_3,\ldots)$, $(y_1,y_2,y_3,\ldots)\in(\mathbb{R}^d)^{\mathbb{N}^*}$. Note that from Hölder inequality the previous conditions (2.3) and (2.4) always hold also for $m=1$.

This value of $m = 1$ is the only one required to prove weak dependence; however higher order moments are needed to derive limit theorems (see [6] and [16]). Assumptions (2.3) and (2.4) are respectively a moment assumption on the variable $F(0, 0, 0, \ldots; \xi_0)$ and a contraction condition on $F(\cdot; \xi_0)$.

THEOREM 2.1. *Assume properties (2.3) and (2.4) for some $m \geq 1$, then there exists a stationary solution $(X_t)_{t\in\mathbb{Z}}$ of Eq. (1.2) such that $\|X_t\|_m < \infty$ that is τ-weakly dependent with*

$$\tau_r \leq C \cdot \left(a^{r/p} + \sum_{k=p}^{\infty} a_k \right) \text{ for some constant } C \text{ only depending on } \mu \text{ and } a.$$

We distinguish two cases

- *If $a_j \leq ce^{-\beta j}$ with $0 < c < \beta$ then $\tau_r \leq Ce^{-\sqrt{(\log\beta - \log c)\beta r}}$.*
- *If $a_j \leq cj^{-\beta}$ with $\beta > 1$ and $0 < c < (\beta-1)/\beta$ then $\tau_r \leq C\left(\frac{\log r}{r}\right)^{\beta}$.*

Remarks..

- As mentioned in the introduction, RSCC are defined through their marginal conditional distributions. We quote here that using such strong conditions as well as an additional irreducibility condition, Iosifescu and Grigorescu (1990) control the coefficient of ϕ-mixing of an RSCC, theorem 2.1.5, page 42 of [19]. This bound has exactly the same form as our bound for τ_r. For this, the authors first control the rate of convergence of such a system to its limit distribution w.r.t. the norm of total variation, in theorem 2.1.3, page 40 of [19]. Here we do not assume any regularity on the inputs (like absolute continuity) which justifies the point that Iosifescu and Grigorescu (1990)'s result cannot be expected here. Andrews (1984) counter-example is a non mixing sequence of some equation like (1.2). We recall here that $(\tau_r)_{r\in\mathbb{Z}}$ ensures most of the limit theorems obtained in the $\phi-$mixing setting, see [8].
- Let $(X_t)_{t\in\mathbb{Z}}$ be a time series only depending on its past, a solution of the Eq. (1.2). Then our conditions (2.3) and (2.4) for $m = 1$ are equivalent to the Dobrushin (1970)'s condition of uniqueness of the stationary measure (if it exists) for $(X_t)_{t\in\mathbb{Z}}$. We prove that in our causal context, Dobrushin (1970)'s condition of uniqueness of a stationary solution also implies its existence. In the case of finite state space, Fernandez and Maillard (2004) denote the same behavior that they call the one-sided Dobrushin condition.

The proof involves a p-Markov approximation stated below. In the following corollary, we propose another way to approximate the stationary measure. Such a constructive result is of independent interest. Set $\tilde{X}_i^{(0)} = 0$ for $1 \le i \le n$ with n fixed, and define recursively for each $k > 0$

$$\tilde{X}_i^{(k)} = F(\tilde{X}_i^{(k-1)}, \dots, \tilde{X}_i^{(k/2)}, 0, \dots; \xi_i). \tag{2.5}$$

COROLLARY 2.1. *Assumes that conditions (2.3) and (2.4) hold, then the distributions of trajectories* $\tilde{\mathbb{X}}_n^{(k)} = (\tilde{X}_1^{(k)}, \dots, \tilde{X}_n^{(k)})'$ *converge to those* $\mathbb{X}_n = (X_1, \dots, X_n)'$ *of the stationary solution* $(X_t)_{t\in\mathbb{Z}}$ *of Eq. (1.2) as* $k \to \infty$. *More precisely,* $\|\tilde{\mathbb{X}}_n^{(k)} - \mathbb{X}_n\|_m \le C \log^{-b} k$ *where* $C > 0$ *and* $b > 0$ *is equal to*
- *If* $a_j \le ce^{-\beta j}$ *with* $0 < c < \beta$ *then* $b = (\log\beta - \log c)/\log 2$.
- *If* $a_j \le cj^{-\beta}$ *with* $\beta > 1$ *and* $0 < c < 1 - 1/\beta$ *then* $b = \beta - 1$.

Depending only on finite number of variables, approximations $\{\tilde{\mathbb{X}}_n^{(k)}\}_{k>0}$ can be simulated. In order to mimick the behavior of a chain with infinite memory we see from Eq. (2.5) that approximations depend on their own past. Remark that $\tilde{\mathbb{X}}_n^{(0)}$ is chosen arbitrarily: thus it does not approximate the trajectory. To forget the first steps error, the approximations $\{\tilde{\mathbb{X}}_n^{(k)}\}_{k>0}$ do not depend on their complete past in Eq. (2.5), contrarily to the chain with infinite memory $(X_t)_{t\in\mathbb{Z}}$.

3. Examples.

This section states precisely some examples. We aim to prove here that many applications may be considered with such class of models.

3.1. *Markov models.*

The first example is the standard and well known p-Markov system

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

Kallenberg (1997) stresses the fact that such equations describe all the p Markov processes.

Non linear autoregressive models have this form with $F(x_1, \dots, x_p; s) = R(x_1, \dots, x_p) + s$; in this case $E = \mathbb{R}^d$ and condition (2.3) follows from $\xi_0 \in \mathbb{L}^m$ while condition (2.4) writes as

$$\|R(y_1, \dots, y_p) - R(x_1, \dots, x_p)\| \le \sum_{j=1}^{p} a_j \|x_j - y_j\|.$$

The fact that the series does not depend of all its past leads to better asymptotic results than those implied by the weak dependence result of Theorem

2.1 (see Duflo in [5] for detailed results).

As soon as $\sum_{j=1}^p a_j = a < 1$, Dedecker and Prieur prove the existence of $\alpha < 1$ such that $\tau_r \le \alpha^r$ in [8]. We improve this rate setting $\tau_r \le a^{r/p}$ (see section 4 for more details).

Example: Random $AR(1)$ models. Such models, solutions of the equation $Y_t = A_t Y_{t-1} + \zeta_t$, occur naturally as iterated random functions; Diaconis and Freedmann (1999) show that fractal behaviors appear for those models. Here $X_t = Y_t$ and $\xi_t = (A_t, \zeta_t)$ with $F(x_1, \dots; (a, \zeta)) = a_1 x_1 + \zeta$. The iid input sequence writes $\xi_t = (A_t, \zeta_t)$ with A_t a $d \times d$-matrix and ζ_t an $\mathbb{R}^d$-valued sequence. This model satisfies $\|F(x;\xi) - F(x';\xi)\|_m \le \|A_0\|_m \|x_1 - x_1'\|$ and the condition (2.4) is satisfied as soon as $\|A_0\|_m < 1$.

3.2. *Random $AR(\infty)$ models.* Let us consider the iid sequence $\{\xi_t = ((A_{j,t})_{j>0}, \zeta_t)\}_{t\in\mathbb{Z}}$ with $A_{j,t}$ a $d \times d$ random matrix for all $j > 0$ and ζ_t an $\mathbb{R}^d$-valued sequence. Here $E = \mathcal{M}_{k,d}(\mathbb{R})^{\mathbb{N}^*} \times \mathbb{R}^d$. An infinite memory extension of random $AR(1)$ models is the solution to the equation

$$X_t = \sum_{j=1}^{\infty} A_{j,t} X_{t-j} + \zeta_t.$$

If $\zeta_0 \in \mathbb{L}^m$ and $\sum_{j=1}^\infty \|A_{j,0}\|_m < 1$, conditions (2.3) and (2.4) hold.

3.3. *Robust bilinear models.* Here $d = 1$ and $E = \mathbb{R}$. Solutions of the equation $X_t = A_t \xi_t + B_t$, where

$$A_t = \sum_{j=1}^{\infty} \alpha_j(X_{t-j}), \qquad B_t = \sum_{j=1}^{\infty} \beta_j(X_{t-j}).$$

extend on the bilinear models in [14] which correspond to linear functions, α_j, β_j. Classical models like ARCH or GARCH processes take this form. Assume $\sum_{j=1}^\infty \mathrm{Lip}\, \alpha_j < 1$, $\sum_{j=1}^\infty \mathrm{Lip}\, \beta_j < 1$ and $\xi_0 \in \mathbb{L}^m$, then conditions (2.3) and (2.4) are satisfied. A simple class of such models is provided with a fixed Lipschitz function h and $\alpha_j(x) = a_j h(x)$ and $\beta_j(x) = a_j h(x)$ for sequences of constants $\{a_j\}_{j>0}$, $\{b_j\}_{j>0}$. We consider bounded approximations of identity such that $h(x) = x \vee M \wedge (-M)$ to introduce robustified versions of the models in [14].

3.4. *$NLARCH(\infty)$ models.* A generalization of $LARCH(\infty)$ models in [15] is given by equation

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

where now $X_t \in \mathbb{R}^d$, ξ_t is a $d \times k$ matrix (here $E = \mathcal{M}_{k,d}(\mathbb{R})$), $\alpha \in \mathbb{R}^k$ and $\alpha_j : \mathbb{R}^d \to \mathbb{R}^k$ are Lipschitz functions; in [15] linear functions $\alpha_j(x) = c_j x$ are considered for $k \times d$ matrices c_j. Assumption (2.4) holds as soon as $\|\xi_0\|_m \sum_{j=1}^{\infty} \operatorname{Lip} \alpha_j < 1$.

3.5. *Models with linear inputs.* Let $f : \mathbb{R}^k \times E \to \mathbb{R}^d$ be measurable an satisfy $\|f(t, \xi_0) - f(s, \xi_0)\|_m \leq L\|t - s\|$ for some finite constant, we consider

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

here c_j denote $k \times d$ matrices. Then the previous relation (2.4) holds with $a_j = L\|c_j\|$. This is a very nonlinear case for which one only needs to produce a function of two variables and a sequence of constants. Kac (1959) used such type of mean field models in statistical mechanics in [21].

3.6. *Threshold models.* Setting $\xi_t = (\zeta_t, \nu_t)$, threshold models may be written as

$$X_t = \sum_{j=1}^{J} 1\!\!1_{\nu_t \in E_j} F_j(X_{t-1}, X_{t-2}, \ldots; \zeta_t)$$

for a measurable partition of $(E_j)_{1 \leq j \leq J}$ of the space of values E' of ν_t. More complicated threshold models are given by a measurable partition $E_1, \ldots, E_J$ of $\mathbb{R}^k \times E'$ and

$$X_t = \sum_{j=1}^{J} 1\!\!1_{\{(G(X_{t-1}, X_{t-2}, \ldots), \nu_t) \in E_j\}} F_j(X_{t-1}, X_{t-2}, \ldots; \zeta_t)$$

for a given function G of the past of the process with value in $\mathbb{R}^k$. It is simple to explicit sufficient conditions for (2.3) and (2.4) to hold.

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

$$X_t = \xi_t K_t + H_t, \tag{3.1}$$

where $K_t = K(X_{t-1}, \ldots)$ and $H_t = H(X_{t-1}, \ldots)$ are both functions of the past values of $(X_{t-1}, X_{t-2}, X_{t-3}, \ldots)$. This case covers several of the previous examples. Using ideas from [14] we prove the existence of marginal densities involving only regularity assumptions on the innovations (and additional conditions on the process (K_t)).

PROPOSITION 3.1 (regularity of affine models). *Assume that innovations $(\xi_t)_{t\in\mathbb{Z}}$ in the model (3.1) admit a common bounded marginal density f_ξ, then if moreover $K \geq 0$, the marginal densities $f_{X_1,\ldots,X_n}$ of $(X_1, \ldots, X_n)$ exist for all fixed $n > 0$ and satisfy for a suitable constant c*

$$\|f_{X_1,\ldots,X_n}\|_\infty \leq c\|f_\xi\|_\infty^n.$$

By integration we obtain the existence and the boundary of the marginal density of X_0 and of the joint densities of the couples $(X_0, X_k)_{k>0}$. Such results are useful to obtain non parametric estimators behavior under weak dependence (see [24] for more details).

Proof of Proposition 3.1. The sequences K_j, H_j may be rewritten as functions of the past of the innovations

$$\begin{aligned} K_j &= k_j(\xi_{j-1}, \ldots, \xi_1, T), \\ H_j &= h_j(\xi_{j-1}, \ldots, \xi_1, T), \end{aligned}$$

where T denotes the past of the innovations. Then we can compute easily that for all continuous and bounded function f

$$\begin{aligned} \mathbb{E}f(X_1, \ldots, X_n) &= \mathbb{E}f((\xi_j K_j + H_j)_{j=1,\ldots,n}) \\ &= \mathbb{E}f((\xi_j k_j(\xi_{j-1}, \ldots, \xi_1, T) + h_j(\xi_{j-1}, \ldots, \xi_1, T))_{j=1,\ldots,n}) \\ &= \int \cdots \int f_\xi(s_1) \cdots f_\xi(s_n) \mathbb{E}f((s_j k_j(s_{j-1}, \ldots, s_1, T) \\ &+ h_j(s_{j-1}, \ldots, s_1, T))_{j=1,\ldots,n}) ds_1 \ldots ds_n. \end{aligned}$$

We change the variables s_j as

$$x_j = s_j k_j(s_{j-1}, \ldots, s_1, T) + h_j(s_{j-1}, \ldots, s_1, T)$$

in order to find the density of $(X_1, \ldots, X_n)$

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathbb{E} \prod_{j=1}^n f_\xi\left(\frac{x_j - h_j(s_{j-1}, \ldots, s_1, T)}{k_j(s_{j-1}, \ldots, s_1, T)}\right) \frac{1}{k_j(s_{j-1}, \ldots, s_1, T)}. \square$$

4. Proofs of the main results. We firstly prove a useful algebraic result, see Lemma 4.1. We then approximate the solution of the Eq. (1.2) by a process with the $p-$Markov process. Existence of the infinite memory chain is obtained with $p \to \infty$ in subsection 2.1. Weak dependence properties are derived from both a sharp control of the coefficient τ for the $p-$Markov approximations (see Lemma 4.4) and the use of coupling techniques. In corollary 2.1 we derive a way to simulate trajectories of our model.

4.1. *An algebraic preliminary.*

Lemma 4.1. *Let $u_0 \geq 0$ and $(u_n)_{n\in\mathbb{Z}}$ be a real sequence such that $|u_n| \leq u_0$ if $n < 0$. Assume that*

$$u_n = \sum_{i=1}^{p} \alpha_i u_{n-i}, \qquad \forall n \geq 0, \tag{4.1}$$

where $(\alpha_1, \dots, \alpha_p)$ are fixed nonnegative numbers with $\alpha = \sum_{i=1}^p \alpha_i$. Then,

$$u_n \leq \alpha^{n/p} u_0, \qquad \forall n \geq 0.$$

Proof. Let us denote by $Q(x) = x^p - \alpha_1 x^{p-1} - \ldots - \alpha_p x^0$ and by ρ the largest modulus of a root for the polynomial Q. We may bound this modulus ρ of the largest zero of Q, this zero is real and non negative as this is proved in ([23], exercice 20, page 106) and

$$\rho \leq \max_{1\leq i\leq p} \left(\frac{\alpha_i}{c_i}\right)^{1/i},$$

for all $c_1, \dots, c_n \geq 0$ such that $c_1 + \cdots + c_n \leq 1$. With the special choice $c_i \leq \alpha_i(1+\epsilon)$ where $1/(1+\epsilon) \geq \alpha$, we obtain $\rho \leq \alpha^{1/p}$.

Set now $U_n^{(p)} = (u_n, \dots, u_{n-p+1})'$ and C_α the companion matrix associated to $Q(x)$. We define a norm on $\mathbb{R}^p$ by $\|(z_1,\dots,z_p)\|_b = \max_{1\leq i\leq p} |z_i b^{i-1}|$, for some $b \in]0,1]$ to be fixed later. Then

$$\left\|U_n^{(p)}\right\|_b = \left\|C_\alpha U_{n-1}^{(p)}\right\|_b.$$

Simple calculations yield

$$\begin{aligned}
\|C_\alpha\|_b &= \sup_{\|z\|_b\leq 1} \|C_\alpha z\|_b \\
&= \sup_{\|z\|_b\leq 1} \left|\sum_{i=1}^{p} \alpha_i b^{i-1} z_i\right| \vee \max_{1\leq i\leq p-1} b^i |z_i|. \\
&\leq \sum_{i=1}^{p} \alpha_i b^{1-i} \vee b.
\end{aligned}$$

Choose now $b = \rho$, then

$$\|C_\alpha\|_\rho \le \rho \le \alpha^{1/p}.$$

We write to conclude

$$u_n \le \left\|U_n^{(p)}\right\|_\rho \le \|C_\alpha\|_\rho^n \left\|U_0^{(p)}\right\|_\rho \le \alpha^{n/p} \left\|U_0^{(p)}\right\|_\rho \le \alpha^{n/p} u_0. \qquad \square$$

4.2. *Markov stationary approximation.* In order to construct a solution to the equation (1.2) we consider, for each fixed $p \ge 0$ and $q > 0$ the p-Markov process $(X_{p,q,t})_{t\ge 0}$ defined by $X_{p,q,t} = 0$ for $t \le -q$ and through the recurrence equation for $t > q$

$$X_{p,q,t} = F(X_{p,q,t-1}, \dots, X_{p,q,t-p}, 0, \dots; \xi_t). \tag{4.2}$$

We first notice that $X_{0,q,t} = F(0, \dots; \xi_t)$ for $t > -q$ is an iid sequences. The Lipschitz condition (2.4) implies

$$\begin{aligned}
\|X_{p,q+1,0} - X_{p,q,0}\|_m &\le \sum_{i=1}^p a_i \|X_{p,q+1,-i} - X_{p,q,-i}\|_m \\
&\le \sum_{i=1}^p a_i \|X_{p,q+1-i,0} - X_{p,q-i,0}\|_m .
\end{aligned}$$

The second inequality derives from that $X_{p,q,-i}$ and $X_{p,q-i,0}$ have the same law from definition for each triplet of positive integers (p, q, i). Let us consider the sequence $v_n = \|X_{p,n+1,0} - X_{p,n,0}\|_m$ for $n \in \mathbb{Z}$. Here $v_n = 0$ if $n < 0$. For $n > 0$

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

Then, $v_n \le u_n$ for any real sequence $(u_n)_{n\in\mathbb{Z}}$ verifying equation (4.1) with $\alpha_i = a_i$, $u_n = 0$ for $n < 0$ and $u_0 = v_0$. Then, using Lemma 4.1, we achieve to the bound

$$v_n \le a^{n/p} v_0 \le a^{n/p} \|X_{p,1,0}\|_m \le a^{n/p} \|F(0, \dots; \xi_t)\|_m \le a^{n/p} \mu_0.$$

Hence, for each p, the sequence $(X_{p,n,0})_{n\in\mathbb{N}}$ has the Cauchy property in $\mathbb{L}^m$; it converges thus in $\mathbb{L}^m$ to some $X_{p,0} \in \mathbb{L}^m$. It is also clear that $X_{p,n,0}$ is measurable w.r.t. the $\sigma-$algebra generated by $\{\xi_t, t \le 0\}$. The $\mathbb{L}^m$-convergence ensures that this is also the case for $X_{p,0}$. Hence we may write $X_{p,0} = H_p(\xi_0, \xi_{-1}, \dots)$ for some measurable function H_p. The same argument for each $t \in \mathbb{Z}$ proves that $X_{p,t} = H_p(\xi_t, \xi_{t-1}, \xi_{t-2}, \dots)$.

Set $\overline{X}_{p,n,t} = (X_{p,n,t}, \ldots, X_{p,n,t-p})$ for each $t \in \mathbb{Z}$. As previously, the $\left(\mathbb{R}^d\right)^{p+1}$-valued sequence $(\overline{X}_{p,n,0})_{n\in\mathbb{N}}$ converges in $\mathbb{L}^m$ (as $n \to \infty$). Its marginals satisfy the relation (4.2). Let now $n \uparrow \infty$, a continuity argument (on F) implies $X_{p,0} = F(X_{p,-1}, \ldots, X_{p,-p}, 0, \ldots; \xi_0)$. For each $t \in \mathbb{Z}$ we apply the same argument to $\overline{X}_{p,n,t}$. Then the sequence $(X_{p,t})_{t\in\mathbb{Z}}$ is a stationary solution of the recurrence equation (4.2) for each $p \geq 0$.
Consider now

$$\mu_p = \|X_{p,t}\|_m, \qquad \mu = \sup_{p\geq 0} \mu_p, \tag{4.3}$$

$$\Delta_{p,t} = \|X_{p+1,t} - X_{p,t}\|_m, \qquad \Delta_p = \sup_{t\in\mathbb{Z}} \Delta_{p,t}, \tag{4.4}$$

note that this definition of μ_p set here for $p > 0$ also extends to $p = 0$ since $X_{0,t} = F(0, \ldots; \xi_t)$ satisfies by definition $\|X_{0,t}\|_m = \mu_0$ from (2.3). Then from (2.4) we first derive

$$\begin{aligned}
\mu_p &\leq \|X_{p,t} - X_{0,t}\|_m + \mu_0, \\
&\leq \sum_{j=1}^{p} a_j \|X_{p,t-j}\|_m + \mu_0, \\
&\leq \mu_p \sum_{j=1}^{p} a_j + \mu_0, \qquad \text{hence} \\
\mu_p &\leq \frac{1}{1-a} \cdot \mu_0.
\end{aligned}$$

We have thus proved the following useful bound:

Lemma 4.2. *Assume properties (2.3) and (2.4) for some $m \geq 1$, then the expression defined by (4.3) satisfies the bound*

$$\mu = \sup_{p\geq 0} \mu_p \leq \frac{\mu_0}{1-a}.$$

We now estimate analogously

$$\begin{aligned}
\Delta_{p,t} &= \|F(X_{p+1,t-1}, \ldots, X_{p+1,t-p-1}, 0, \ldots; \xi_t) \\
&\qquad\qquad - F(X_{p,t-1}, \ldots, X_{p,t-p}, 0, \ldots; \xi_t)\|_m \\
&\leq \sum_{j=1}^{p} a_j \|X_{p+1,t-j} - X_{p,t-j}\|_m + a_{p+1} \|X_{p+1,t-p-1}\|_m \\
&\leq \sum_{j=1}^{p} a_j \Delta_{p,t-j} + a_{p+1} \|X_{p+1,t-p-1}\|_m
\end{aligned}$$

The previous equation implies $\Delta_p \leq \dfrac{a_{p+1}}{1-a} \cdot \mu$, thus with lemma 4.2 we derive the following lemma

LEMMA 4.3. *Assume properties (2.3) and (2.4) for some $m \geq 1$, then the expression defined by (4.3) satisfies the bound*

$$\Delta_p \leq a_{p+1} \cdot \frac{\mu_0}{(1-a)^2}.$$

4.3. *Proof of existence in Theorem 2.1.* Quote first that lemma 4.3 implies that $X_{p,t} \to X_t$ in $\mathbb{L}^m$ since this space is complete. The continuity of F ensures that X_t is a solution of equation (1.2). Furthermore, as a limit in $\mathbb{L}^m$ of stationary process, X_t is also stationary (in law). Finally, as a limit in $\mathbb{L}^m$ of a process admitting a moment of order m, $\|X_t\|_m < \infty$. $\square$

4.4. *Proof of weak dependence in Theorem 2.1.* Previous assumptions with $m = 1$ are enough for this section. We first need:

LEMMA 4.4. *Assume that (2.4) and (2.3) hold with $m = 1$, then the Markov chain $(X_{p,t})_t$ in Eq. (4.2) is weakly dependent with*

$$\tau_{p,r} \leq a^{r/p}.$$

Proof of the lemma 4.4. We follow here the proof of Duflo [5], as Dedecker and Prieur in [8]. We improve their result by using the lemma 4.1. As quoted in section (2), we use coupling in order to evaluate the τ-coefficient. Let $(\xi'_t)_t$ be a process distributed as $(\xi_t)_t$ and independent of the innovation. We define the process $(X^*_{p,t})_t$ as

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

Using the approximation of the section 4.2, we equivalently find a sequence of measurable variables w.r.t. the $\sigma-$algebra generated by $\xi'_t, t \leq 0$ denoted $(X^*_{p,n,0})_{n \in \mathbb{N}}$ such that it converges in $\mathbb{L}^m$ to $X^*_{p,0} \in \mathbb{L}^m$. The $\mathbb{L}^m$-convergence ensures this is also the case for $X_{p,0}$. Then, by definition of $\xi'_t, t \leq 0$, $X^*_{p,0}$ is independent of $X_{p,0}$. Due to the coupling property of the coefficient τ, we obtain $\tau_{p,r} \leq \|X_{p,r} - X^*_{p,r}\|_1$.

It is easy to check that assumption (2.4) leads to

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

Denoting $w_r = \|X_{p,r} - X^*_{p,r}\|_1$ for $r \in \mathbb{Z}$, we use again the lemma 4.2 and we obtain

$$\begin{aligned}
\tau_{p,r} &\leq w_r \\
&\leq a^{r/p} w_0 \\
&\leq 2\mu_p a^{r/p} \\
&\leq 2\frac{\|F(0,\dots;\xi_0)\|_1}{1-a} \cdot a^{r/p}
\end{aligned}$$

Indeed if $m = 1$, $\|F(0,\dots;\xi_0)\|_1 = \mu_0$. $\square$

Let us now define the process $(X^*_t)_t$ by

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

We remark that $(X^*_t)_t$ is also a stationary chain with infinite memory. Lemma 4.3 gives us

$$\|X_r - X_{p,r}\|_1 \leq \sum_{k=p}^{\infty} \Delta_k \leq C \sum_{k=p}^{\infty} a_{k+1}.$$

The same bound also holds for the quantity $\|X^*_r - X^*_{p,r}\|_1$. For each integer p

$$\begin{aligned}
\tau_r &\leq \|X_r - X^*_r\|_1, \\
&\leq \|X_r - X_{p,r}\|_1 + \|X_{p,r} - X^*_{p,r}\|_1 + \|X^*_r - X^*_{p,r}\|_1, \\
&\leq A\left(a^{r/p} + \sum_{k=p}^{\infty} a_{k+1}\right).
\end{aligned}$$

We now choose p such that both terms in this sum have the same decay rate. If $a_j \leq ce^{-\beta j}$ then we choose $p = \sqrt{\alpha r/\beta}$ where $e^{-\alpha} = a < 1$, and then we found the bound $\tau_r \leq C\left(e^{-\sqrt{\alpha\beta r}}\right)$ for a suitable constant $C > 0$. If $a_j \leq cj^{-\beta}$ we find $\tau_r \leq C\left(\left(\frac{\log r}{r}\right)^{\beta}\right)$. $\square$

4.5. *Proof of corollary 2.1.* We define recursively the process $\left(\tilde{X}_0^{(n)}\right)_{n\in\mathbb{N}}$ by $\tilde{X}_0^{(0)} = 0$ and by the relation

$$\tilde{X}_0^{(n)} = F\left(\tilde{X}_0^{(n-1)}, \dots, \tilde{X}_0^{(n-p_n)}, 0, \dots; \xi_0\right),$$

here, the sequence $(p_n)_{n\in\mathbb{N}}$ will be precisely set later. Then, the Lipschitz assumption (2.4) leads to

$$\left\|\tilde{X}_0^{(n+1)} - X_0\right\|_m \le \sum_{i=1}^{p_n} a_i \left\|\tilde{X}_0^{(n+1-i)} - X_0\right\|_m + \sum_{i>p_n} a_i \left\|X_0\right\|_m, \qquad n \ge 1$$

Fix some integer $N > 0$. Then, denoting by $v_n = \left\|\tilde{X}_0^{(n+1)} - X_0\right\|_m$ for $0 \le n \le N$, this sequence satisfies the recursion

$$v_n \ \le \ \sum_{i=1}^{p_n} a_i v_{n-i} + \varepsilon_{p_n},$$

with $\varepsilon_{p_n} = \|X_0\|_m \sum_{i>p_n} a_i \to 0$ as soon as $p_n \to \infty$ with n. Let us choose a sequence $q_n \le p_n$ and denote $e_N = \max\{\varepsilon_k; k \ge q_N\}$. Then for all $n' \le n \le N$ such that $q_N \le p_{n'}$ we have

$$v_n \ \le \ \sum_{i=1}^{p_N} a_i v_{n-i} + e_N.$$

Then, $v_n \le u_{n-n'}$ for any real sequence $(u_n)_{n\in\mathbb{Z}}$ verifying equation

$$u_n = \sum_{i=1}^{p_N} a_i u_{n-i} + e_N,$$

$u_n = 0$ for $n < 0$ and $u_0 = v_{n'}$ for $i = 0$. The solution of this equation is the sum of the solution of the homogeneous equation (4.1) with $\alpha_i = a_i$ and a special solution. $e_N/(1-\sum_{i=1}^{p_N} a_i)$ is a solution and we obtain for $n' \le n \le N$:

$$v_n \le u_{n-n'} \le a^{(n-n')/p_N} u_0 + \frac{e_N}{1-a} \le a^{(n-n')/p_N} v_{n'} + \frac{e_N}{1-a}.$$

We choose $n = N$ to get $v_N \le a^{(N-n')/p_N} v_{n'} + e_N/(1-a)$.
Taking $q_n = [qn]$ and $p_n = -[-pn]$ for $0 < q < p$ we obtain $n' \le Nq/p$ and

$$v_N \le a^{1/p-q/p^2} v_{Nq/p} + e_N/(1-a).$$

The special choices $q = 1/4$ and $p = 1/2$ lead to $v_N \le a v_{N/2} + e_N/(1-a)$. Then, for the integers n satisfying $2^{n-1} < N \le 2^n$, we derive

$$\begin{aligned} v_N &\le a^n v_1 + \frac{1}{1-a} \sum_{i=0}^{n-1} a^i e_{N/2^i}, \\ &\le N^{\log a/\log 2} \mu_0 + \frac{\log N}{1-a} \max_{0\le i\le n-1} a^i e_{N/2^i}. \end{aligned}$$

We check precisely two special cases

- $a_i = (a/(1+a))^i$, then $\max\limits_{0 \le i \le n-1} a^i e_{N/2^i} \le \|X_0\|_m N^{\log a/\log 2}$.
- $a_i = i^{-\beta}$, here $\max\limits_{0 \le i \le n-1} a^i e_{N/2^i} \le \|X_0\|_m N^{1-\beta}$. $\square$

REFERENCES

[1] Andrews, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21**, 4, 930–934. MRMR766830 (86e:60027) MR0766830

[2] Berbee, H. (1987). Chains with infinite connections: uniqueness and Markov representation. *Probab. Theory Related Fields* **76**, 2, 243–253. MRMR906777 (89c:60052) MR0906777

[3] Bühlmann, P. and Wyner, A. J. (1999). Variable length Markov chains. *Ann. Statist.* **27**, 2, 480–513. MRMR1714720 (2000j:62123) MR1714720

[4] Comets, F., Fernández, R., and Ferrari, P. A. (2002). Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12**, 3, 921–943. MRMR1925446 (2003f:60068) MR1925446

[5] Duflo, M. (1997). *Random iterative models*. Applications of Mathematics (New York), Vol. **34**. Springer-Verlag, Berlin. MRMR1485774 (98m:62239) MR1485774

[6] Dedecker, J. and Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Process. Appl.* **106**, 1, 63–80. MRMR1983043 (2004c:60111) MR1983043

[7] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., Prieur, C. (2005) Weak dependence, examples and applications. monograph in preparation.

[8] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132**, 2, 203–236. MRMR2199291 MR2199291

[9] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 1, 45–76 (electronic). MRMR1669737 (2000c:60102) MR1669737

[10] Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory Probab. Applic.* 15, 458-486.

[11] Dobrushin, R. L. and Kusuoka, S. (1993). *Statistical mechanics and fractals*. Lecture Notes in Mathematics, Vol. **1567**. Springer-Verlag, Berlin. MRMR1295841 (96b:82018) MR1295841

[12] Doukhan, P. (1994). *Mixing*. Lecture Notes in Statistics, Vol. **85**. Springer-Verlag, New York. MRMR1312160 (96b:60090) MR1312160

[13] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84**, 2, 313–342. MRMR1719345 (2001j:60033) MR1719345

[14] Doukhan, P, Madre, H. and Rosenbaum, M. (2005). ARCH type bilinear weakly dependent models. *Submitted for publication, available at http://www.crest.fr/pageperso/doukhan/publications.htm*

[15] Doukhan, P., Teyssière, G. and Winant, P. (2006). A LARCH(∞) Vector Valued Process. *In Dependence in Probability and Statistics, Lecture Notes in Statistics,* vol. 187. Bertail, P; Doukhan, P; Soulier, Ph; (Eds.), Springer.

[16] Doukhan, P. and Wintenberger, O. (2005). Invariance principle for new weakly dependent stationary models under sharp moment assumptions. *Submitted for publication, available at http://www.crest.fr/pageperso/doukhan/publications.htm*

[17] Fernández, R. and Maillard, G. (2004). Chains with complete connections and one-dimensional Gibbs measures. *Electron. J. Probab. 9*, no. 6, 145–176 (electronic). MRMR2041831 (2005a:60047) MR2041831

[18] Georgii, H.-O. (1988). *Gibbs measures and phase transitions.* de Gruyter Studies in Mathematics, Vol. **9**. Walter de Gruyter and Co., Berlin. MRMR956646 (89k:82010) MR0956646

[19] Iosifescu, M. and Grigorescu, Ş. (1990). *Dependence with complete connections and its applications.* Cambridge Tracts in Mathematics, Vol. **96**. Cambridge University Press, Cambridge. MRMR1070097 (91j:60098) MR1070097

[20] Iosifescu, M., , and Theodorescu, R. (1969). *Random processes and learning.* Springer-Verlag, New York. MRMR0293704 (45 #2781) MR0293704

[21] Kac, M. (1959). *Probability and related topics in physical sciences.* Interscience Publishers, London-New York, MRMR0102849 (21 #1635),

[22] Kallenberg, O. (1997). *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York. MRMR1464694 (99e:60001) MR1464694

[23] Polya, G. and Szegö, G. (1978). Problems and theorems in Analysis 1. *Classics in Mathematics, Springer Verlag,* Berlin.

[24] Ragache, N. and Wintenberger, O. (2006). Convergence rates for density estimators of weakly dependent time series. *In Dependence in Probability and Statistics, Lecture Notes in Statistics*, vol. 187. Bertail, P; Doukhan, P; Soulier, Ph; (Eds.), Springer.

[25] Rosenblatt, M. (1985). *Stationary sequences and random fields.* Birkhäuser Boston Inc., Boston, MA. MRMR885090 (88c:60077) MR0885090

[26] Wu, W. B. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic Process. Appl.* **115**, 6, 939–958. MRMR2138809 (2005m:62176) MR2138809

LS-CREST,
Laboratoire de Statistique,
Timbre J340,
3, avenue Pierre Larousse,
92240 MALAKOFF
e-mail: doukhan@ensae.fr

SAMOS-MATISSE
(Statistique Appliquée et MOdélisation Stochastique)
Centre d'Economie de la Sorbonne
Universit Paris 1 - Panthon-Sorbonne, CNRS
90, rue de Tolbiac
75634-PARIS CEDEX 13
e-mail: olivier.wintenberger@univ-paris1.fr