On fixed gain recursive estimators with discontinuity in the parameters

Citation for published version:
Chau, HN, Kumar, C, Rásonyi, M & Sabanis, S 2019, 'On fixed gain recursive estimators with discontinuity in the parameters', *ESAIM: Probability and Statistics*, vol. 23, pp. 217-244.
https://doi.org/10.1051/ps/2018019

Digital Object Identifier (DOI):
10.1051/ps/2018019

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
ESAIM: Probability and Statistics

On fixed gain recursive estimators with discontinuity in the parameters

Huy N. Chau* Chaman Kumar† Miklós Rásonyi‡

Sotirios Sabanis§

October 16, 2018

Abstract

In this paper we estimate the expected tracking error of a fixed gain stochastic approximation scheme. The underlying process is not assumed Markovian, a mixing condition is required instead. Furthermore, the updating function may be discontinuous in the parameter.

MSC 2010 subject classification: Primary: 62L20; secondary: 93E15, 93E35

Keywords: adaptive control, stochastic gradient, fixed gain, recursive estimators, parameter discontinuity, mixing processes, non-Markovian dynamics

1 Introduction

Let $\mathbb{N} := \{0, 1, 2, \ldots\}$. We are interested in stochastic approximation procedures where a parameter estimate θ_t, $t \in \mathbb{N}$ is updated by a recursion of the form

$$\theta_{t+1} = \theta_t + \gamma_{t+1} H(\theta_t, X_{t+1}),\ t \in \mathbb{N}, \tag{1}$$

starting from some guess θ_0. Here X_t is a stationary signal, γ_t is a sequence of real numbers and $H(\cdot,\cdot)$ is a given functional. The most common choices are $\gamma_t = 1/t$ (decreasing gain) and $\gamma_t := \lambda$ (fixed gain). The former family of procedures is aimed to converge to θ^* with $G(\theta^*) = 0$ where $G(\theta) := EH(\theta, X_t)$. The latter type of procedures is supposed to "track" θ^*, even when the system dynamics is (slowly) changing.

In most of the related literature the error analysis of (1) was carried out only in the case where H is (Lipschitz-)continuous in θ. This restrictive hypothesis fails to accommodate discontinuous procedures which are common in practice, e.g. the signed regressor, signed error and sign-sign algorithms (see [3], [7], [8]) or the Kohonen algorithm (see [27, 1]). Recently, the decreasing gain case was investigated in [9] for controlled Markov chains and the procedure (1) was shown to converge

*MTA Alfréd Rényi Institute of Mathematics, Budapest. E-mail: chau@renyi.mta.hu

†Indian Institute of Technology, Roorkee. E-mail: c.kumarfma@iitr.ac.in

‡MTA Alfréd Rényi Institute of Mathematics, Budapest. E-mail: rasonyi@renyi.mta.hu

§University of Edinburgh. E-mail: S.Sabanis@ed.ac.uk

almost surely under appropriate assumptions, without requiring continuity of H. We refer to [9] for a review of the relevant literature and for examples.

The purpose of the present article is an exploration of the case where X_t has possibly non-Markovian dynamics. We consider fixed gain procedures and weaken continuity of H to continuity in the sense of conditional expectations, see (6) below, compare also to condition **H4** in [9].

We follow the methodology of the papers [19, 14, 17] which are based on the concept of L-mixing, coming from [12]. Our arguments work under a modification of the original definition of L-mixing, see Section 2. We furthermore assume a certain asymptotic forgetting property, see Assumption 3.4. We manage to estimate the tracking error for (1), see our main result, Theorem 3.6 in Section 3.

At this point we would like to make comparisons with another important reference, [28], where no Markovian or continuity assumptions were made, certain averaging properties of the driving process were required instead. It follows from Subsection 4.2 of [28] that almost sure convergence of a decreasing gain procedure can be guaranteed under the α-mixing property of the driving process, see e.g. [6] about various mixing concepts. It seems that establishing the L-mixing property is often relatively simple while α-mixing is rather stringent and difficult to prove. In addition, our present work provides explicit estimates for the error. See Section 4 for examples illustrating the scope of Theorem 3.6.

Section 5 reports simulations showing that the theoretical estimate is in accordance with numerical results. Proofs for Sections 2 and 3 are relegated to Section 6.

2 L-mixing and conditional L-mixing

Estimates for the error of stochastic approximation schemes like (1) can be proved under various ergodicity assumptions on the driving process. It is demonstrated in [14] and [17] that the concept of L-mixing (see its definition below in the present section) is sufficiently strong for this purpose. An appealing feature of L-mixing is that it can easily be applied in non-Markovian contexts as well, see Section 4.

It turns out, however, that for discontinuous updating functions H the arguments of [14, 17] break down. To tackle discontinuities, we introduce a new concept of mixing here, which is of interest on its own right.

Throughout this paper we are working on a probability space $(\Omega, \mathcal{F}, P)$ that is equipped with a discrete-time filtration $\mathcal{F}_n$, $n \in \mathbb{N}$ as well as with a decreasing sequence of sigma-fields $\mathcal{F}_n^+$, $n \in \mathbb{N}$ such that $\mathcal{F}_n$ is independent of $\mathcal{F}_n^+$, for all n.

Expectation of a random variable X will be denoted by EX. For any $m \geq 1$, for any $\mathbb{R}^m$-valued random variable X and for any $1 \leq p < \infty$, let us set $\|X\|_p := \sqrt[p]{E|X|^p}$. We denote by L^p the set of X satisfying $\|X\|_p < \infty$. The indicator function of a set A will be denoted by 1_A.

We now present the class of L-mixing processes which were introduced in [12]. This concept proved to be extremely useful in solving certain hard problems of system identification, see e.g. [15, 16, 13, 18, 30].

Fix an integer $N \geq 1$ and let $D \subset \mathbb{R}^N$ be a set of parameters. A measurable function $X : \mathbb{N} \times D \times \Omega \to \mathbb{R}^m$ is called a random field. We will drop dependence on $\omega \in \Omega$ and use the notation $X_t(\theta)$, $t \in \mathbb{N}$, $\theta \in D$.

For any $r \geq 1$, a random field $X_t(\theta)$, $t \in \mathbb{N}$, $\theta \in D$ is called *bounded in L^r* if

$$M_r(X) := \sup_{\theta \in D} \sup_{t \in \mathbb{N}} \|X_t(\theta)\|_r < \infty. \quad (2)$$

For an L^r-bounded $X_t(\theta)$, define also the quantities

$$\gamma_r(\tau, X) := \sup_{\theta \in D} \sup_{t \geq \tau} \|X_t(\theta) - E[X_t(\theta)|\mathcal{F}^+_{t-\tau}]\|_r, \ \tau \geq 1,$$

and

$$\Gamma_r(X) := \sum_{\tau=1}^{\infty} \gamma_r(\tau, X). \quad (3)$$

For some $r \geq 1$, a random field $X_t(\theta)$ is called *uniformly L-mixing of order r* (ULM-r) if it is bounded in L^r; for all $\theta \in D$, $X_t(\theta)$, $t \in \mathbb{N}$ is adapted to $\mathcal{F}_t$, $t \in \mathbb{N}$; and $\Gamma_r(X) < \infty$. Here uniformity refers to the parameter θ. Furthermore, $X_t(\theta)$ is called *uniformly L-mixing* if it is uniformly L-mixing of order r for all $r \geq 1$.

In the case of a single stochastic process (which corresponds to the case where the parameter set D is a singleton) we apply the terminology "L-mixing process of order r" and "L-mixing process".

Remark 2.1. The L-mixing property shows remarkable stability under various operations, this is why it proved to be a versatile tool in the analysis of stochastic systems, see [14, 17, 15, 16, 13, 18, 30]. If F is a Lipschitz function and $X_t(\theta)$ is ULM-r then $F(X_t(\theta))$ is also ULM-r, by (33) in Lemma 6.1 below. Actually, if F is such that $|F(x) - F(y)| \leq K(1 + |x|^k + |y|^k)|x - y|$ for all $x, y \in \mathbb{R}$ with some $k, K > 0$ then $F(X_t(\theta))$ is uniformly L-mixing whenever $X_t(\theta)$ is, see Proposition 2.4 of [30]. Stable linear filters also preserve the L-mixing property, see [12]. Proving that $F(X_t(\theta))$ is L-mixing for discontinuous F is more delicate, see Section 4 for helpful techniques.

Other mixing conditions could alternatively be used. Some of these are inherited by arbitrary measurable functions of the respective processes (e.g. ϕ-mixing, see Section 7.2 of [6]). However, they are considerably difficult to verify while L-mixing (and its conditional version to be defined below) is relatively simple to check, see also the related remarks on page 2129 of [18].

Recall that, for any family Z_i, $i \in I$ of real-valued random variables, $\operatorname{ess.sup}_{i \in I} Z_i$ denotes a random variable that is an almost sure upper bound for each Z_i and it is a.s. smaller than or equal to any other such bound. Such an object is known to exist, independently of the cardinality of I, and it is a.s. unique, see e.g. Proposition VI.1.1. of [29].

Now we define conditional L-mixing, inspired by (2) and (3). Let $X_t(\theta)$, $t \in \mathbb{N}$, $\theta \in D$ be a random field bounded in L^r for some $r \geq 1$ and define, for each $n \in \mathbb{N}$,

$$\begin{aligned}
M^n_r(X) &:= \operatorname{ess\,sup}_{\theta \in D} \sup_{t \in \mathbb{N}} E^{1/r}[|X_{n+t}(\theta)|^r \big| \mathcal{F}_n], \\
\gamma^n_r(\tau, X) &:= \operatorname{ess\,sup}_{\theta \in D} \sup_{t \geq \tau} E^{1/r}[|X_{n+t}(\theta) - E[X_{n+t}(\theta)|\mathcal{F}^+_{n+t-\tau} \vee \mathcal{F}_n]|^r \big| \mathcal{F}_n], \ \tau \geq 1, \\
\Gamma^n_r(X) &:= \sum_{\tau=1}^{\infty} \gamma^n_r(\tau, X).
\end{aligned}$$

For some $s, r \geq 1$, we call $X_t(\theta)$, $t \in \mathbb{N}$, $\theta \in D$ uniformly *conditionally* L-mixing of order (r, s) (abbreviation: UCLM-(r, s)) if it is L^r-bounded; $X_t(\theta)$, $t \in \mathbb{N}$

is adapted to $\mathcal{F}_t$, $t \in \mathbb{N}$ for all $\theta \in D$ and the sequences $M^n_r(X)$, $\Gamma^n_r(X)$, $n \in \mathbb{N}$ are bounded in L^s. When the UCLM-(r,s) property holds for all $r,s \geq 1$ then we simply say that the random field is uniformly conditionally L-mixing. In the case of stochastic processes (when D is a singleton) the terminology "conditionally L-mixing process of order (r,s)" (respectively, conditionally L-mixing process) will be used.

Remark 2.2. Note that if $\mathcal{F}_0$ is trivial and $X_t(\theta)$ is UCLM-$(r,1)$ then it is also ULM-r. Indeed, in that case

$$M_r(X) = M^0_r(X), \quad \Gamma_r(X) = \Gamma^0_r(X).$$

For non-trivial $\mathcal{F}_0$, however, no such implication holds.

Remark 2.3. If F is a Lipschitz function and $X_t(\theta)$ is UCLM-$(r,1)$ then $F(X_t(\theta))$ is also UCLM-$(r,1)$, by Lemma 6.1 below. Conditional versions of the arguments in Lemma 6.2 show that if $X_t(\theta)$ is UCLM-$(rp,1)$ and $Y_t(\theta)$ is UCLM-$(rq,1)$ (where $1/p + 1/q = 1$) then

$$M^n_r(XY) \leq M^n_{rp}(X)M^n_{rq}(Y), \tag{4}$$

$$\Gamma^n_r(XY) \leq 2M^n_{rp}(X)\Gamma^n_{rq}(Y) + 2\Gamma^n_{rp}(X)M^n_{rq}(Y). \tag{5}$$

We now present another concept, a surrogate for continuity in $\theta \in D$. We say that the random field $X_t(\theta) \in L^1$, $t \in \mathbb{N}$, $\theta \in D$ satisfies the *conditional Lipschitz-continuity* (CLC) property if there is a deterministic $K > 0$ such that, for all $\theta_1, \theta_2 \in D$ and for all $n \in \mathbb{N}$,

$$E\left[|X_{n+1}(\theta_1) - X_{n+1}(\theta_2)| \,\middle|\, \mathcal{F}_n\right] \leq K|\theta_1 - \theta_2|, \text{ a.s.} \tag{6}$$

Pathwise discontinuities of $\theta \to X_n(\theta)$ can often be smoothed out and (6) can be verified by imposing some conditions on the one-step conditional distribution of X_{n+1} given $\mathcal{F}_n$, see Assumption 4.3 and Lemma 4.7 below.

Remark 2.4. We comment on the differences between condition **H4** of [9] and our CLC property. Assume that X is stationary and Markovian. On one hand, **H4** of [9] stipulates that, for $\delta > 0$

$$\sup_{\theta \in D_c} E\left[\sup_{\theta' \in D_c,\ |\theta - \theta'| \leq \delta} |H(\theta, X_1) - H(\theta', X_1)|\right] \leq K\delta^\alpha \tag{7}$$

for any compact $D_c \subset D$ with some $K > 0$ (that may depend on D_c) and with some $0 < \alpha \leq 1$ (independent of D_c). On the other hand, CLC is equivalent to

$$\sup_{\theta, \theta' \in D,\ |\theta - \theta'| \leq \delta} E\left[\left|H(\theta, X_1) - H(\theta', X_1)\right| \,\middle|\, X_0 = x\right] \leq K\delta. \tag{8}$$

for Law(X_0)-almost every x. Clearly, (7) allows Hölder-continuity (i.e. $\alpha < 1$) while (8) requires Lipschitz-continuity. In the case $\alpha = 1$ (7) is not comparable to CLC though both express a kind of "continuity in the average".

The main results of our paper require a specific structure for the sigma-algebras which facilitates to deduce properties of conditional L-mixing processes from those of "unconditional" ones. More precisely, we rely on the crucial Doob-type inequality in Theorem 2.5 below. This could probably be proved for arbitrary sigma-algebras but only at the price of redoing all the tricky arguments of [12] in a more difficult context. We refrain from this since Theorem 2.5 can accommodate most models of practical importance. Let $\mathbb{Z}$ denote the set of integers.

Theorem 2.5. *Fix $r > 2$, $n \in \mathbb{N}$. Assume that, for all $t \in \mathbb{N}$, $\mathcal{F}_t = \sigma(\varepsilon_j,\ j \in \mathbb{N},\ j \le t)$, $\mathcal{F}_t^+ := \sigma(\varepsilon_j,\ j > t)$ for some i.i.d. sequence ε_j, $j \in \mathbb{Z}$ with values in some Polish space $\mathcal{X}$. Let W_t, $t \in \mathbb{N}$ be a conditionally L-mixing process of order $(r,1)$, satisfying $E[W_t|\mathcal{F}_n] = 0$ a.s. for all $t \ge n$. Let $m > n$ and let b_t, $n < t \le m$ be deterministic numbers. Then we have*

$$E^{1/r}\left[\max_{n<t\le m}\left|\sum_{s=n+1}^{t} b_s W_s\right|^r \Big|\mathcal{F}_n\right] \le C_r \left(\sum_{s=n+1}^{m} b_s^2\right)^{1/2} \sqrt{M_r^n(W)\Gamma_r^n(W)}, \qquad (9)$$

almost surely, where C_r is a deterministic constant depending only on r but independent of n, m.

The proof is reported in Section 6.

3 Fixed gain stochastic approximation

Let $N \ge 1$ be an integer and let $\mathbb{R}^N$ be the Euclidean space with norm $|x| := \sqrt{\sum_{i=1}^N x_i^2}$, $x \in \mathbb{R}^N$. Let $D \subset \mathbb{R}^N$ be a bounded (nonempty) open set representing possible system parameters. Let $H : D \times \mathbb{R}^m \to \mathbb{R}^N$ be a bounded measurable function. We assume throughout this section that for all $t \in \mathbb{N}$, $\mathcal{F}_t = \sigma(\varepsilon_j,\ j \in \mathbb{N},\ j \le t)$, $\mathcal{F}_t^+ := \sigma(\varepsilon_j,\ j > t)$ for some i.i.d. sequence ε_j, $j \in \mathbb{Z}$ with values in some Polish space $\mathcal{X}$, in particular the condition on the sigma algebras in the statement of Theorem 2.5 holds.

Let

$$X_t := g(\varepsilon_t, \varepsilon_{t-1}, \dots),\ t \in \mathbb{N}, \qquad (10)$$

with some fixed measurable function $g : \mathcal{X}^{-\mathbb{N}} \to \mathbb{R}^m$. Clearly, X is a (strongly) stationary $\mathbb{R}^m$-valued process, see Lemma 10.1 of [24].

Remark 3.1. We remark that, in the present setting, the CLC property holds if, for all $\theta_1, \theta_2 \in D$,

$$E\left[|H(\theta_1, X_1) - H(\theta_2, X_1)| \Big|\mathcal{F}_0\right] \le K|\theta_1 - \theta_2|, \text{ a.s.},$$

due to the fact that the law of $(X_{k+1}, \varepsilon_k, \varepsilon_{k-1}, \dots)$ is the same as that of $(X_1, \varepsilon_0, \varepsilon_{-1}, \dots)$, for all $k \in \mathbb{Z}$.

Define $G(\theta) := EH(\theta, X_0)$. Note that, by stationarity of X, $G(\theta) = EH(\theta, X_t)$ for all $t \in \mathbb{N}$. We need some stability hypotheses formulated in terms of an ordinary differential equation related to G.

Assumption 3.2. *On D, the function G is twice continuously differentiable and bounded, together with its first and second derivatives.*

Fix $\lambda > 0$. Under Assumption 3.2, the equation

$$\dot{y}_t = \lambda G(y_t), \quad y_s = \xi, \tag{11}$$

has a unique solution for each $s \geq 0$ and $\xi \in D$, on some (finite or infinite) interval $[s, v(s,\xi))$ with $v(s,\xi) > s$. We will denote this solution by $y(t,s,\xi)$, $t \in [s, v(s,\xi))$. Let $D_1 \subset D$ such that for all $\xi \in D_1$ we have $y(t,0,\xi) \in D$ for any $t \geq 0$. We denote

$$\phi(D_1) = \{u \in D : u = y(t,0,\xi), \text{ for some } t \geq 0, \ \xi \in D_1\}.$$

The ε-neighbourhood of a set D_1 is denoted by $S(D_1, \varepsilon)$, i.e.

$$S(D_1, \varepsilon) = \{u \in \mathbb{R}^N : |u - \theta| < \varepsilon \text{ for some } \theta \in D_1\}.$$

We remark that, under Assumption 3.2, the function $y(t,s,\xi)$ is continuously differentiable in ξ.

Notice that all the above observations would be true under weaker hypotheses than those of Assumption 3.2. However, the proof of Lemma 6.5 below requires the full force of Assumption 3.2, see [17].

Assumption 3.3. *There exist open sets*

$$\emptyset \neq D_\xi \subset D_y \subset D_\theta \subset D_{\overline{y}} \subset D$$

such that $\phi(D_\xi) \subset D_y$, $S(D_y, d) \subset D_\theta$ *for some* $d > 0$ *and* $\phi(D_\theta) \subset D_{\overline{y}}$, $S(D_{\overline{y}}, d') \subset D$ *for some* $d' > 0$. *The ordinary differential equation (11) is exponentially asymptotically stable with respect to initial perturbations, i.e. there exist* $C^* > 0, \alpha > 0$ *such that, for each* λ *sufficiently small, for all* $0 \leq s \leq t$, $\xi \in D$

$$\left| \frac{\partial}{\partial \xi} y(t,s,\xi) \right| \leq C^* e^{-\lambda\alpha(t-s)}. \tag{12}$$

We furthermore assume that there is $\theta^* \in D$ *such that*

$$G(\theta^*) = 0. \tag{13}$$

It follows from $\phi(D_\xi) \subset D_y$ and (12) that θ^* actually lies in the closure of D_y and that there is only one θ^* satisfying (13).

While Assumptions 3.2, 3.3 pertained to a deterministic equation, our next hypothesis is of a stochastic nature.

Assumption 3.4. *For all* $n \in \mathbb{N}$,

$$E\left[\sup_{\vartheta \in D} \sum_{k=n}^{\infty} \left| E[H(\vartheta, X_{k+1}) | \mathcal{F}_n] - G(\vartheta) \right| \right] < \infty \tag{14}$$

Remark 3.5. Assumption 3.4 expresses a certain kind of "forgetting": for k large, $E[H(\vartheta, X_{k+1})|\mathcal{F}_n]$ is close to $G(\vartheta) = EH(\theta, X_{k+1})|_{\theta=\vartheta}$ in L^1, uniformly in ϑ and the convergence is fast enough so that the sum in (14) is finite. In other words, this is again a kind of mixing property.

In certain cases, the validity of Assumption 3.4 indeed follows from L-mixing. Let X_t, $t \in \mathbb{N}$ be L-mixing of order 1 and let $x \to H(\theta, x)$ be Lipschitz-continuous with a Lipschitz constant $L^{\dagger}$ that is independent of θ. We claim that Assumption 3.4 holds under these conditions. Indeed, for every $\vartheta \in D$

$$\sum_{k=n}^{\infty} E\left|E\left[H(\vartheta, X_{k+1})|\mathcal{F}_n\right] - E[H(\theta, X_{k+1})]|_{\theta=\vartheta}\right| \leq$$

$$\sum_{k=n}^{\infty}\left|E\left[H(\vartheta, X_{k+1})|\mathcal{F}_n\right] - E\left[H(\vartheta, E[X_{k+1}|\mathcal{F}_n^+])|\mathcal{F}_n\right]\right| +$$

$$\sum_{k=n}^{\infty}\left|E\left[H(\theta, E[X_{k+1}|\mathcal{F}_n^+])\right]|_{\theta=\vartheta} - E[H(\theta, X_{k+1})]|_{\theta=\vartheta}\right| \leq$$

$$2L^{\dagger}\sum_{k=n}^{\infty}\left|X_{k+1} - E[X_{k+1}|\mathcal{F}_n^+]\right|$$

noting that

$$E\left[H(\vartheta, E[X_{k+1}|\mathcal{F}_n^+])|\mathcal{F}_n\right] = E\left[H(\theta, E[X_{k+1}|\mathcal{F}_n^+])\right]|_{\theta=\vartheta},$$

by independence of $\mathcal{F}_n$ and $\mathcal{F}_n^+$. Hence

$$E\left[\sup_{\vartheta\in D}\sum_{k=n}^{\infty}\left|E[H(\vartheta, X_{k+1})|\mathcal{F}_n] - G(\vartheta)\right|\right] \leq 2L^{\dagger}\Gamma_1(X) < \infty.$$

Assumption 3.4 can also be verified in certain cases where H is discontinuous, see Section 4.

We now state the main result of our article.

Theorem 3.6. *Let $H(\theta, X_t)$ be UCLM-$(r,1)$ for some $r > 2$, satisfying the CLC property (see (6) above). Let Assumptions 3.2, 3.3 and 3.4 be in force. For some $\xi \in D_\xi$, define the recursive procedure*

$$\theta_0 := \xi, \quad \theta_{t+1} = \theta_t + \lambda H(\theta_t, X_{t+1}), \tag{15}$$

with some $\lambda > 0$. Define also its "averaged" version,

$$z_0 := \xi, \quad z_{t+1} = z_t + \lambda G(z_t). \tag{16}$$

Let d, d' in Assumption 3.3 be large enough and let λ be small enough. Then $\theta_t, z_t \in D_\theta$ for all t and there is a constant C, independent of $t \in \mathbb{N}$ and of λ, such that

$$E\left|\theta_t - z_t\right| \leq C\lambda^{1/2},\ t \in \mathbb{N}.$$

An important consequence of the main theorem is provided as follows.

Corollary 3.7. *Under the conditions of Theorem 3.6, there is $t_0(\lambda) \in \mathbb{N}$ such that*

$$E\left|\theta_t - \theta^*\right| \leq C\lambda^{1/2},\ t \geq t_0(\lambda).$$

Furthermore, $t_0(\lambda) \leq C^\circ \ln(1/\lambda)/\lambda$ for some $C^\circ > 0$.

The proofs of Theorem 3.6 and Corollary 3.7 are postponed to Section 6.

Remark 3.8. Our current investigations were motivated by [17] where not only the random field $H(\theta, X_t)$ was assumed L-mixing, but also its "derivative field"

$$\frac{H(\theta_1, X_t) - H(\theta_2, X_t)}{\theta_1 - \theta_2}, \ t \in \mathbb{N}, \ \theta_1, \theta_2 \in D, \ \theta_1 \neq \theta_2. \tag{17}$$

As shown in Section 3 of [12], the latter hypothesis necessarily implies the continuity (in θ) of $H(\theta, X_t)$. For our purposes such an assumption is thus too strong. We are able to drop continuity at the price of modifying the L-mixing concept, as explained in Section 2 above.

We point out that our results complement those of [17] even in the case where H is Lipschitz-continuous (in that case the CLC property of our paper obviously holds). In [17], the derivative field (17) was assumed to be L-mixing. In the present paper we do not need this hypothesis (but we assume conditional L-mixing of order $(r, 1)$ for some $r > 2$ instead of L-mixing).

4 Examples

The present section serves to illustrate the power of Theorem 3.6 above by exhibiting processes X_t and functions H to which that theorem applies.

The (conditional) L-mixing property can be verified for arbitrary bounded measurable functionals of Markov processes with the Doeblin condition (see [20]) and this could probably be extended to a larger family of Markov processes using ideas of [2] or [22]. We prefer not to review the corresponding methods here but to present some non-Markovian examples because they demonstrate better the advantages of our approach over the existing literature.

In Subsection 4.1 linear processes (see e.g. Subsection 3.2 of [21]) with polynomial autocorrelation decay are considered, while Subsection 4.2 presents a class of Markov chains in a random environment with contractive properties.

4.1 Causal linear processes

Assumption 4.1. *Let ε_j, $j \in \mathbb{Z}$ be a sequence of independent, identically distributed real-valued random variables such that $E|\varepsilon_0|^\zeta < \infty$ for some $\zeta \geq 2$ and $E\varepsilon_0 = 0$. We set $\mathcal{F}_n = \sigma(\varepsilon_i, i \leq n)$, and $\mathcal{F}_n^+ = \sigma(\varepsilon_i, i > n)$ for each $n \in \mathbb{Z}$. Let us define the process*

$$X_t := \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{18}$$

where $a_j \in \mathbb{R}$, $j \in \mathbb{N}$. We assume $a_0 \neq 0$ and

$$|a_j| \leq C_1 (j+1)^{-\beta}, \ j \in \mathbb{N},$$

for some constants $C_1 > 0$ and $\beta > 1/2$.

Note that the series (18) converges a.s. (by Kolmogorov's theorem, see e.g. Chapter 4 of [24]). As a warm-up, we now check the conditional L-mixing property for X.

Lemma 4.2. *Let Assumption 4.1 be in force. If $\beta > 3/2$ then the process X_t, $t \in \mathbb{N}$ is conditionally L-mixing of order $(\zeta, 1)$.*

Proof. We have, for $t \in \mathbb{N}$,

$$
\begin{aligned}
E^{1/\zeta}[|X_t|^\zeta|\mathcal{F}_0] &\leq E^{1/\zeta}\left[2^{\zeta-1}|\sum_{j=0}^{t-1} a_j \varepsilon_{t-j}|^\zeta \Big|\mathcal{F}_0\right] + E^{1/\zeta}\left[2^{\zeta-1}|\sum_{j=t}^{\infty} a_j \varepsilon_{t-j}|^\zeta \Big|\mathcal{F}_0\right] \\
&\leq 2^{\frac{\zeta-1}{\zeta}} \sum_{j=0}^{t-1} |a_j| \|\varepsilon_{t-j}\|_\zeta + 2^{\frac{\zeta-1}{\zeta}} \sum_{j=t}^{\infty} |a_j \varepsilon_{t-j}|
\end{aligned}
$$

using the simple inequality $(x+y)^\zeta \leq 2^{\zeta-1}(x^\zeta + y^\zeta)$, $x, y \geq 0$; properties of the norm $\|\cdot\|_\zeta$; independence of ε_j, $j \geq 1$ from $\mathcal{F}_0$ and $\mathcal{F}_0$-measurability of ε_j, $j \leq 0$. Hence

$$
\begin{aligned}
E^{1/\zeta}[|X_t|^\zeta|\mathcal{F}_0] &\leq 2^{\frac{\zeta-1}{\zeta}} \|\varepsilon_0\|_\zeta \sum_{j=0}^{\infty} |a_j| + 2^{\frac{\zeta-1}{\zeta}} \sum_{j=0}^{\infty} C_1 (t+j+1)^{-\beta} |\varepsilon_{-j}| \\
&\leq 2^{\frac{\zeta-1}{\zeta}} \|\varepsilon_0\|_\zeta \sum_{j=0}^{\infty} |a_j| + 2^{\frac{\zeta-1}{\zeta}} \sum_{j=0}^{\infty} C_1 (j+1)^{-\beta} |\varepsilon_{-j}| \\
&\leq C_2 \left[1 + \sum_{j=0}^{\infty} |\varepsilon_{-j}|(j+1)^{-\beta}\right],
\end{aligned}
$$

for some $C_2 > 0$. Note that the latter bound is independent of t. Similar estimates prove that, for all $n \geq 0$,

$$
M_\zeta^n(X) \leq C_2 \left[1 + \sum_{j=0}^{\infty} |\varepsilon_{n-j}|(j+1)^{-\beta}\right].
$$

The right-hand side has the same law for all n and it is in L^1 since $\beta > 1$. This implies that the sequence $M_\zeta^n(X)$, $n \in \mathbb{N}$ is bounded in L^1.

For $1 \leq m$ and for any $t \in \mathbb{Z}$, define

$$
X_{t,m}^+ := \sum_{j=0}^{m-1} a_j \varepsilon_{t-j},
$$

and, for $t \geq m$, let

$$
X_{t,m}^\circ := X_{t,m}^+ + \sum_{j=t}^{\infty} a_j \varepsilon_{t-j}.
$$

Notice that $E[X_t|\mathcal{F}_{t-m}^+ \vee \mathcal{F}_0] = X_{t,m}^\circ$ and, by independence of ε_j, $j \geq 1$ from $\mathcal{F}_0$,

$$
\begin{aligned}
E[|X_t - X_{t,m}^\circ|^\zeta \big|\mathcal{F}_0] = \left\|\sum_{j=m}^{t-1} a_j \varepsilon_{t-j}\right\|_\zeta^\zeta &\leq \qquad (19) \\
C_3 E\left(\sum_{j=m}^{t-1} a_j^2 \varepsilon_{t-j}^2\right)^{\zeta/2} &\leq \\
C_3 \left(\sum_{j=m}^{t-1} \|a_j^2 \varepsilon_{t-j}^2\|_{\zeta/2}\right)^{\zeta/2} &\leq \\
C_3 \left(C_1^2 \sum_{j=m}^{\infty} (1+j)^{-2\beta} \|\varepsilon_0\|_\zeta^2\right)^{\zeta/2} &\leq \left(C_3' m^{-2\beta+1}\right)^{\zeta/2},
\end{aligned}
$$

with some constants $C_3, C_3' > 0$, using the Marczinkiewicz-Zygmund inequality. Define $b_m := \sqrt{C_3'} m^{-\beta+1/2}$. An analogous estimate gives $\gamma_\zeta^n(m,X) \le b_m$ for all $n \in \mathbb{N}$. Since $\sum_{m=1}^\infty b_m < \infty$ by $\beta > 3/2$, $\Gamma_\zeta^n(X)$ is actually bounded by a constant, uniformly in n. □

We also need in the sequel that the law of the driving noise is smooth enough. This is formulated in terms of the characteristic function ϕ of ε_0.

Assumption 4.3. *We require that*

$$\int_{\mathbb{R}} |\phi(u)|\, du < \infty. \tag{20}$$

Remark 4.4. Assumption 4.3 implies the existence of a (continuous and bounded) density f for the law of ε_0 (with respect to the Lebesgue measure). Indeed, f is the inverse Fourier transform of ϕ:

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(u) e^{-iux}\, du, \ x \in \mathbb{R}.$$

Conversely, if the law of ε_0 has a twice continuously differentiable density f such that f', f'' are integrable over $\mathbb{R}$ then (20) holds. The latter observation follows by standard Fourier-analytic arguments.

Lemma 4.5. *Let Assumptions 4.1 and 4.3 be in force. Then the law of X_0 (resp. $X_{0,m}^+$) has a density f_∞ (resp. f_m) with respect to the Lebesgue measure. Moreover, there is a constant $\tilde{K} > 0$ such that*

$$\sup_{m \in \mathbb{N} \cup \{\infty\}} \sup_{x \in \mathbb{R}} f_m(x) \le \tilde{K}.$$

Proof. Denote by ϕ_m the characteristic function of $X_{0,m}^+$. Since $|\phi(u)| \le 1$ for all u, we see that

$$|\phi_m(u)| = \left| \prod_{j=0}^{m-1} \phi(a_j u) \right| \le |\phi(a_0 u)|, \tag{21}$$

which implies, by applying an inverse Fourier transform, the existence of f_m and the estimate

$$|f_m(x)| \le \frac{1}{2\pi} \int_{\mathbb{R}} |\phi(a_0 u)|\, du =: \tilde{K} < \infty, \text{ for all } x \in \mathbb{R},\ m \ge 1,$$

by Assumption 4.3. As $X_{0,m}^+$ tends to X_0 in probability when $m \to \infty$, $\phi_m(u)$ tends to $\phi_\infty(u)$ for all u, where ϕ_∞ is the characteristic function of X_0. The integrable bound (21) is uniform in m, so f_∞ exists and the dominated convergence theorem implies that $f_m(x)$ tends to $f_\infty(x)$, for all $x \in \mathbb{R}$. The result follows. □

Let $D \subset \mathbb{R}^N$ be a bounded open set. In the sequel we consider functionals of the form

$$H(\theta, x) := \sum_{j=1}^{M} g_j(\theta, x) 1_{\{x \in I_j(\theta)\}},\ x \in \mathbb{R},\ \theta \in D, \tag{22}$$

where the g_j are bounded and Lipschitz-continuous functions (jointly in the two variables) and the intervals $I_j(\theta)$ are of the form $(-\infty, h_j(\theta))$, $(h_j(\theta), \infty)$ or $(h_j^1(\theta), h_j^2(\theta))$ with $h_j, h_j^1, h_j^2 : D \to \mathbb{R}$ Lipschitz-continuous functions.

Remark 4.6. The intervals $I_j(\theta)$ can also be closed or half-closed and the results below remain valid, this is clear from the proofs. In the one-dimensional case, the signed regressor, signed error, sign-sign and Kohonen algorithms all have an updating function of the form (22), see [7], [8], [27], [1]. For simplicity, we only treat the one-dimensional setting (i.e. $x \in \mathbb{R}$) in the present paper but we allow D to be multidimensional.

Lemma 4.7. *Let Assumptions 4.1 and 4.3 be in force. Then a random field $H(\theta, X_t)$, $t \in \mathbb{N}$, $\theta \in D$ as in* (22) *satisfies the CLC property* (6).

Proof. It suffices to consider $H(\theta, X_1) = g(\theta, X_1)1_{\{X_1 \in I(\theta)\}}$ with g Lipschitz-continuous, bounded and I of the form $(-\infty, h(\theta))$, $(h(\theta), \infty)$ or $(h^1(\theta), h^2(\theta))$ with h, h^1, h^2 Lipschitz. We only prove the first case, the other cases being similar. Recall also Remark 3.1.

Denoting by C_4 a Lipschitz-constant for g and by C_5 an upper bound for $|g|$, we get the estimate

$$\begin{aligned} |H(\theta_1, X_1) - H(\theta_2, X_1)| &\leq \\ |1_{\{X_1 < h(\theta_1)\}} g(\theta_1, X_1) - 1_{\{X_1 < h(\theta_1)\}} g(\theta_2, X_1)| &+ \\ |1_{\{X_1 < h(\theta_1)\}} g(\theta_2, X_1) - 1_{\{X_1 < h(\theta_2)\}} g(\theta_2, X_1)| &\leq \\ C_4|\theta_1 - \theta_2| + C_5|1_{\{X_1 < h(\theta_1)\}} - 1_{\{X_1 < h(\theta_2)\}}| &\leq \\ C_4|\theta_1 - \theta_2| + C_5\left(1_{\{X_1 \in [h(\theta_1), h(\theta_2))\}} + 1_{\{X_1 \in [h(\theta_2), h(\theta_1))\}}\right). & \end{aligned}$$

We may and will assume $h(\theta_1) < h(\theta_2)$. It suffices to prove that

$$P\left(X_1 \in [h(\theta_1), h(\theta_2)) | \mathcal{F}_0\right) \leq C_6|\theta_1 - \theta_2|$$

with a suitable $C_6 > 0$. Noting that the density of $a_0\varepsilon_1$ is $x \to (1/|a_0|)f(x/a_0)$, we have

$$\begin{aligned} P\left(X_1 \in [h(\theta_1), h(\theta_2)) | \mathcal{F}_0\right) &= \int_{h(\theta_1) - \sum_{j=1}^{\infty} a_j \varepsilon_{1-j}}^{h(\theta_2) - \sum_{j=1}^{\infty} a_j \varepsilon_{1-j}} \frac{1}{|a_0|} f(x/a_0)\, dx \leq \\ \frac{1}{|a_0|} K_0 |h(\theta_1) - h(\theta_2)| &\leq \frac{1}{|a_0|} K_0 C_7 |\theta_1 - \theta_2|, \end{aligned}$$

where K_0 is an upper bound for f (see Remark 4.4) and C_7 is a Lipschitz constant for h. This completes the proof. □

Theorem 4.8. *Let Assumptions 4.1 and 4.3 be in force. Let H be of the form specified in* (22). *Let $\zeta \geq 2r$ and let β satisfy $\beta > 4r + 1/2$. Then the random field $H(\theta, X_t)$, $t \in \mathbb{N}$, $\theta \in D$ is UCLM-$(r, 1)$.*

Proof. We may and will assume

$$H(\theta, X_t) = g(\theta, X_t)1_{\{X_t \in I(\theta)\}}$$

with some bounded Lipschitz function g and with some interval $I(\theta)$ of the type as in (22). As H is bounded, $M_r^n(H)$, $n \in \mathbb{N}$ is trivially a bounded sequence in L^1.

In view of (4), (5), it suffices to establish that $1_{\{X_t \in I(\theta)\}}$, $t \in \mathbb{N}$ is UCLM-$(2r, 1)$ (since $g(\theta, X_t)$, $1_{\{X_t \in I(\theta)\}}$ are bounded and $g(\theta, X_t)$ is UCLM-$(2r, 1)$, by $\zeta \geq 2r$, Lemma 4.2 and Remark 2.3). We show this for $I(\theta) = (-\infty, h(\theta))$ with h Lipschitz-continuous as other types of intervals can be handled similarly.

As

$$\mathrm{Law}(X_{t+n},\ t \in \mathbb{N},\ \varepsilon_{n-j},\ j \in \mathbb{N}) = \mathrm{Law}(X_t,\ t \in \mathbb{N},\ \varepsilon_{-j},\ j \in \mathbb{N})$$

for all $n \in \mathbb{N}$, we may reduce the proof to estimations for the case $n := 0$. Let us start with

$$\begin{aligned} \left| 1_{\{X^\circ_{t,m} < h(\theta)\}} - 1_{\{X_t < h(\theta)\}} \right| &= \\ 1_{\{X^\circ_{t,m} < h(\theta), X_t \geq h(\theta)\}} + 1_{\{X_t < h(\theta), X^\circ_{t,m} \geq h(\theta)\}} &\leq \\ 1_{\{X_t \in (h(\theta) - \eta_m, h(\theta) + \eta_m)\}} + 1_{\{|X^\circ_{t,m} - X_t| \geq \eta_m\}}, \end{aligned}$$

for all $\eta_m > 0$. We will choose a suitable η_m later. Using Lemma 4.5 and the conditional Markov inequality we obtain

$$\begin{aligned} E\left[|1_{\{X^\circ_{t,m} < h(\theta)\}} - 1_{\{X_t < h(\theta)\}}|^{2r} | \mathcal{F}_0 \right] &\leq \\ C_8 P(X_t \in (h(\theta) - \eta_m, h(\theta) + \eta_m) | \mathcal{F}_0) &+ \\ C_8 P(|X^\circ_{t,m} - X_t| \geq \eta_m | \mathcal{F}_0) &\leq \\ 2C_8 \tilde{K} \eta_m + C_8 E\left[|X_t - X^\circ_{t,m}| \middle| \mathcal{F}_0 \right] / \eta_m, \end{aligned} \tag{23}$$

with some constant C_8, noting that powers of indicators are themselves indicators and that the conditional density of X_t with respect to $\mathcal{F}_0$ is $x \to f_t(x - \sum_{j=t}^{\infty} a_j \varepsilon_{t-j})$ and the latter is $\leq \tilde{K}$ by Lemma 4.5. Using (19), the second term in (23) is bounded by $C_8 \sqrt{C_3'} m^{-\beta+1/2} / \eta_m$ hence it is reasonable to choose $\eta_m := 1/m^{(\beta-1/2)/2}$, which leads to

$$E^{1/2r}\left[|1_{\{X^\circ_{t,m} < h(\theta)\}} - 1_{\{X_t < h(\theta)\}}|^{2r} \middle| \mathcal{F}_0 \right] \leq C_9 / m^{(\beta-1/2)/(4r)},$$

with some $C_9 > 0$. Notice that $X^\circ_{t,m}$ is $\mathcal{F}^+_{t-m} \vee \mathcal{F}_0$-measurable. Lemma 6.1 implies that $\gamma_{2r}(X, m) \leq 2C_9 / m^{(\beta-1/2)/(4r)}$. As $(\beta - 1/2)/(4r) > 1$ by our hypotheses, we obtain the UCLM-$(2r, 1)$ property for $1_{\{X_t \in I(\theta)\}}$. □

Remark 4.9. When ε_0 has moments of all orders then one can reduce the lower bound $4r + 1/2$ for β in Theorem 4.8 to $r + 1/2$. Indeed, in this case $g(\theta, X_t)$ is UCLM-$(q, 1)$ for arbitrarily large q by Lemma 4.2 and Remark 2.3 so it suffices to show the UCLM-$(r', 1)$ property for $1_{\{X_t \in I(\theta)\}}$ for some $r' > r$ that can be arbitrarily close to r (and not for $r' = 2r$ as in Theorem 4.8). The estimate of the above proof can be improved to

$$\begin{aligned} E\left[|1_{\{X^\circ_{t,m} < h(\theta)\}} - 1_{\{X_t < h(\theta)\}}|^{r'} | \mathcal{F}_0 \right] &\leq \\ 2C_8 \tilde{K} \eta_m + C_8 E\left[|X_t - X^\circ_{t,m}|^q \middle| \mathcal{F}_0 \right] / \eta_m^q, \end{aligned}$$

for arbitrarily large q. Choosing $\eta_m := 1/m^{[q(\beta-1/2)]/(q+1)}$, we arrive at

$$E^{1/r'}\left[|1_{\{X^\circ_{t,m} < h(\theta)\}} - 1_{\{X_t < h(\theta)\}}|^{r'} \middle| \mathcal{F}_0 \right] \leq C_9 / m^{[q(\beta-1/2)]/[(q+1)r']}.$$

Let $\beta > r + 1/2$. If $r' > r$ is chosen close enough to r and q is chosen large enough then $[q(\beta - 1/2)]/[(q+1)r'] > 1$ which shows the UCLM-$(r, 1)$ property for $H(\theta, X_t)$.

Lemma 4.10. *Let Assumptions 4.1 and 4.3 be in force, let $\beta > 3/2$. Then, for all $n \in \mathbb{N}$,*

$$E\left[\sum_{k=n}^{\infty} \sup_{\vartheta\in D} \left| E\left[H(\vartheta, X_{k+1})|\mathcal{F}_n\right] - G(\vartheta)\right|\right] \leq C_{10}$$

with some fixed $C_{10} < \infty$. That is, Assumption 3.4 holds.

Proof. We need to estimate

$$\sum_{k=n}^{\infty}\left[\left|E\left[g(\vartheta, X_{k+1})1_{\{X_{k+1}<h(\vartheta)\}}|\mathcal{F}_n\right] - E[g(\theta, X_{k+1})1_{\{X_{k+1}<h(\theta)\}}]|_{\theta=\vartheta}\right|\right],$$

where $h : D \to \mathbb{R}$ is Lipschitz-continuous and g is a bounded, Lipschitz-continuous function with a bound C_{11} for $|g|$ and with Lipschitz constant C_{12}. It suffices to prove

$$E\left[\sup_{\vartheta\in D}\sum_{k=1}^{\infty}\left|E\left[g(\vartheta, X_0)1_{\{X_0<h(\vartheta)\}}|\mathcal{F}_{-k}\right] - E[g(\theta, X_0)1_{\{X_0<h(\theta)\}}]|_{\theta=\vartheta}\right|\right] < \infty,$$

since the law of $(X_0, \varepsilon_{-k}, \varepsilon_{-k-1}, \ldots)$ equals that of $(X_{n+k}, \varepsilon_n, \varepsilon_{n-1}, \ldots)$, for all $k \geq 1$, $n \in \mathbb{Z}$. We can estimate a given term in the above series as follows:

$$\begin{aligned}
\left|E\left[g(\vartheta, X_0)1_{\{X_0<h(\vartheta)\}}|\mathcal{F}_{-k}\right] - E[g(\theta, X_0)1_{\{X_0<h(\theta)\}}]|_{\theta=\vartheta}\right| &\leq \\
\left|E\left[g(\vartheta, X_0)1_{\{X_0<h(\vartheta)\}}|\mathcal{F}_{-k}\right] - E[g(\vartheta, X^+_{0,k})1_{\{X^+_{0,k}<h(\vartheta)\}}|\mathcal{F}_{-k}]\right| &+ \\
\left|E[g(\theta, X^+_{0,k})1_{\{X^+_{0,k}<h(\theta)\}}]|_{\theta=\vartheta} - E[g(\theta, X_0)1_{\{X_0<h(\theta)\}}]|_{\theta=\vartheta}\right| &\leq \\
C_{12}E\left[|X_0 - X^+_{0,k}|\middle|\mathcal{F}_{-k}\right] + C_{11}E\left[|1_{\{X^+_{0,k}<h(\vartheta)\}} - 1_{\{X_0<h(\vartheta)\}}|\middle|\mathcal{F}_{-k}\right] &+ \\
C_{12}|X_0 - X^+_{0,k}| + C_{11}E\left[|1_{\{X^+_{0,k}<h(\theta)\}} - 1_{\{X_0<h(\theta)\}}|\right]|_{\theta=\vartheta} & \qquad (24)
\end{aligned}$$

noting that $E[g(\theta, X^+_{0,k})1_{\{X^+_{0,k}<h(\theta)\}}]|_{\theta=\vartheta} = E[g(\vartheta, X^+_{0,k})1_{\{X^+_{0,k}<h(\vartheta)\}}|\mathcal{F}_{-k}]$. The first and third terms on the right-hand side of (24) are equal and they are $\leq C_{13}k^{-\beta+1/2}$ with some $C_{13} > 0$, by the proof of Lemma 4.2, hence their sum (when k goes from 1 to infinity) is finite. The expression in the second term of (24) can be estimated as

$$\begin{aligned}
E\left[|1_{\{X^+_{0,k}<h(\vartheta)\}} - 1_{\{X_0<h(\vartheta)\}}|\middle|\mathcal{F}_{-k}\right] &\leq \\
P\left(X_0 \in \left(h(\vartheta) - \sum_{j=k}^{\infty} a_j\varepsilon_{-j}, h(\vartheta) + \sum_{j=k}^{\infty} a_j\varepsilon_{-j}\right)|\mathcal{F}_{-k}\right) &\leq \\
2\tilde{K}\left|\sum_{j=k}^{\infty} a_j\varepsilon_{-j}\right|, & \qquad (25)
\end{aligned}$$

noting that the conditional density of X_0 with respect to $\mathcal{F}_{-k}$ is $x \to f_k\left(x - \sum_{j=k}^{\infty} a_j\varepsilon_{-j}\right)$ and this is bounded by $\tilde{K}$, using Lemma 4.5. Since (25) is independent of ϑ, a similar estimate guarantees that

$$E\left[|1_{\{X^+_{0,k}<h(\theta)\}} - 1_{\{X_0<h(\theta)\}}|\right]|_{\theta=\vartheta} \leq 2\tilde{K}\left|\sum_{j=k}^{\infty} a_j\varepsilon_{-j}\right|.$$

Note that the upper estimates obtained so far do not depend on ϑ. It follows that, even taking supremum in $\vartheta \in D$, the expectations of the second and fourth terms on the right-hand side of (24) are both $\leq C_{14}k^{(-\beta+1/2)}$ with some $C_{14} > 0$. As $\beta > 3/2$, the infinite sum of these terms is finite, too, finishing the proof of the present lemma. □

Assumption 4.11. *Let f satisfy*

$$|f(x)| \leq \widehat{C}e^{-\widehat{\delta}|x|} \text{ for all } x \in \mathbb{R}, \tag{26}$$

with some $\widehat{C}, \widehat{\delta} > 0$

Assumption 4.12. *Let*

$$\int_{\mathbb{R}} u^2 |\phi_u| \, du < \infty \tag{27}$$

hold.

Remark 4.13. Clearly, Assumption 4.11 implies that ε_0 has finite moments of all orders. Note also that Assumption 4.12 implies Assumption 4.3.

Remark 4.14. If f is four times continuously differentiable such that f', f'', f''', f'''' are integrable then (27) holds, compare to Remark 4.4 above.

Lemma 4.15. *Let Assumptions 4.1, 4.11 and 4.12 hold. Let the functions g_i, h_i^1, h_i^2, h_i of (22) be twice continuously differentiable with bounded first and second derivatives. Then $G(\theta) := EH(\theta, X_0)$ is bounded and twice continuously differentiable with bounded first and second derivatives, i.e. Assumption 3.2 holds.*

Proof. We may and will assume that

$$H(\theta, x) = 1_{\{x<h(\theta)\}} g(\theta, x)$$

with h Lipschitz-continuous, g bounded and Lipschitz-continuous. G is bounded since g is. We proceed to establish its differentiability and the boundedness of its derivatives.

Recall that

$$\phi_\infty(u) = \prod_{j=0}^{\infty} \phi(a_j u), \ u \in \mathbb{R},$$

where ϕ_∞ is the characteristic function of X_0 and the product converges pointwise. Since $|\phi(u)| \leq 1$ for all u, (27) implies that

$$\int_{\mathbb{R}} u^2 |\phi_\infty(u)| \, du < \infty.$$

Clearly, this implies $\int_{\mathbb{R}} |u\phi_\infty(u)| \, du < \infty$ and $\int_{\mathbb{R}} |\phi_\infty(u)| \, du < \infty$ as well (since ϕ_∞ is bounded, being a Fourier transform). Now one can directly show, using the inverse Fourier transform, that f_∞, the density of the law of X_0, is twice continuously differentiable.

Inequality (26) implies that ϕ has a complex analytic extension in a strip around $\mathbb{R}$. Since the sequence a_j, $j \in \mathbb{N}$ is bounded, there is even a strip such that $u \to \phi(a_j u)$ is analytic in it, for all $j \in \mathbb{N}$, thus ϕ_∞ is also analytic there. Then so are $-iu\phi_\infty(u)$ and $-u^2\phi_\infty(u)$. These being integrable, we get that their inverse Fourier transforms, f'_∞ and f''_∞, satisfy

$$|f'_\infty(x)| + |f''_\infty(x)| \le \tilde{C} e^{-\tilde{\delta}|x|} \text{ for all } x \in \mathbb{R}, \tag{28}$$

with some $\tilde{C}, \tilde{\delta} > 0$, see e.g. Theorem 11.9.3 of [25]. In particular, f'_∞, f''_∞ are integrable.

For notational simplicity we consider only the case $N = 1$, i.e. $D \subset \mathbb{R}$. Using the change of variable $y = x - h(\theta)$, we see that

$$EH(\theta, X_0) = \int_{\mathbb{R}} g(\theta, x) 1_{\{x<h(\theta)\}} f_\infty(x)\, dx = \int_{-\infty}^{0} g(\theta, y + h(\theta)) f_\infty(y + h(\theta))\, dy.$$

We calculate $\partial_\theta g(\theta, y + h(\theta)) f_\infty(y + h(\theta))$:

$$[\partial_1 g(\theta, y+h(\theta)) + \partial_2 g(\theta, y+h(\theta)) h'(\theta)] f_\infty(y+h(\theta)) + g(\theta, y+h(\theta)) f'_\infty(y+h(\theta)) h'(\theta),$$

where ∂_1 (resp. ∂_2) denote differentiation with respect to the first (resp. second) variable. As f_∞ (resp. f'_∞) satisfy (26) (resp. (28)) and $g, \partial_1 g, \partial_2 g, h'$ are bounded, the dominated convergence theorem implies that

$$\begin{aligned} \partial_\theta EH(\theta, X_0) &= \\ \int_{-\infty}^{0} [\partial_1 g(\theta, y + h(\theta)) + \partial_2 g(\theta, y + h(\theta)) h'(\theta)] f_\infty(y + h(\theta))\, dy &+ \\ \int_{-\infty}^{0} g(\theta, y + h(\theta)) f'_\infty(y + h(\theta)) h'(\theta)\, dy &= \\ \int_{\mathbb{R}} 1_{\{x<h(\theta)\}} [\partial_1 g(\theta, x) + \partial_2 g(\theta, x) h'(\theta)] f_\infty(x)\, dx &+ \\ \int_{\mathbb{R}} 1_{\{x<h(\theta)\}} g(\theta, x) f'_\infty(x) h'(\theta)\, dx, & \end{aligned}$$

where both integrals are clearly bounded in θ. Similar calculations involving the second derivatives of g, h, f_∞ show that $\partial^2_\theta EH(\theta, X_0)$ exists and it is bounded in θ. □

The following corollary summarizes our findings in the present section.

Corollary 4.16. *Let H be of the form* (22) *such that g_j, h_j, h^1_j, h^2_j are twice continuously differentiable with bounded first and second derivatives. Let Assumptions 4.1, 4.11 and 4.12 hold and assume $\beta > 5/2$. Then Theorem 3.6 applies to the random field $H(\theta, X_t)$, $t \in \mathbb{N}$, $\theta \in D$, provided that Assumption 3.3 holds.*

Proof. Recalling 4.13 and 4.9, this corollary follows from the results of the present section. □

Assumptions 4.11 and 4.12 apply, in particular, when ε_0 is Gaussian. There does not seem to be a general condition guaranteeing the validity of Assumption 3.3: this needs checking in every concrete application of Theorem 3.6.

4.2 Markov chains in a random environment

If $\beta \leq 5/2$ in the setting of Subsection 4.1 above then Corollary 4.16 cannot be established with our methods. Hence X_t cannot be a "long memory processes" in the sense of [21]. In this subsection we show that it is nonetheless possible to apply Theorem 3.6 to important classes of random fields that are driven by a long memory process, see Example 4.18 below.

Let ε_t^1, $t \in \mathbb{Z}$, ε_t^2, $t \in \mathbb{Z}$ be i.i.d. real-valued sequences, independent of each other.

Assumption 4.17. *We denote $\chi_t := (\varepsilon_{j+t}^1)_{j\in\mathbb{Z}}$, for each $t \in \mathbb{Z}$. Let $F : \mathbb{R}^{\mathbb{Z}} \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ be a measurable function such that, for all $w \in \mathbb{R}^{\mathbb{Z}}$, $s \in \mathbb{R}$,*

$$|F(w,s,z_1) - F(w,s,z_2)| \leq \rho|z_1 - z_2|,$$

for all $z_1, z_2 \in \mathbb{R}^m$ with some $0 < \rho < 1$. Furthermore, there is $x \in \mathbb{R}^m$ such that for all $w \in \mathbb{R}^{\mathbb{Z}}$ and for all $s \in \mathbb{R}$,

$$|F(w,s,x) - x| \leq C(1 + |s|)$$

and $E|\varepsilon_0^2|^r < \infty$ for some $r > 2$.

Fix $x \in \mathbb{R}$ as in Assumption 4.17 and define, for all $t \in \mathbb{Z}$, $\tilde{X}_0^t := x$, and for $j \geq 0$,

$$\tilde{X}_{j+1}^t := F(\chi_t, \varepsilon_t^2, \cdot) \circ F(\chi_{t-1}, \varepsilon_{t-1}^2, \cdot) \circ \cdots F(\chi_{t-j}, \varepsilon_{t-j}^2, \cdot)(x).$$

Standard arguments (such as Proposition 5.1 of [5]) show that $\tilde{X}_j^t$ converges almost surely as $j \to \infty$. Define

$$X_t := \lim_{j\to\infty} \tilde{X}_j^t.$$

Then X_t, $t \in \mathbb{Z}$ is clearly a stationary process, satisfying

$$X_{t+1} = F(\chi_{t+1}, \varepsilon_{t+1}^2, X_t),\ t \in \mathbb{Z}.$$

When freezing the values of ε_t^1, $t \in \mathbb{Z}$, the X_t defined above is an (inhomogeneous) Markov chain driven by the noise sequence ε_t^2, $t \in \mathbb{Z}$. Hence X_t is a Markov chain in a random environment (the latter is driven by ε_t^1, $t \in \mathbb{Z}$).

Example 4.18. Let ε_i^2, $i \in \mathbb{Z}$ be i.i.d. with $E|\varepsilon_0^2|^r < \infty$ for some $r > 2$. Let $E\varepsilon_0^1 = 0$, $E(\varepsilon_0^1)^2 < \infty$. Let $Y_t := \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}^1$, $t \in \mathbb{Z}$ for some a_j, $j \in \mathbb{N}$ with $\sum_{j=0}^{\infty} a_j^2 < \infty$. The series converges almost surely. Let $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ be bounded measurable and fix $-1 < \kappa, \rho < 1$. The construction sketched above provides the existence of a process X_t satisfying

$$X_{t+1} = \kappa X_t + \rho e^{h_1(Y_{t+1})} h_2(\varepsilon_{t+1}^1) + \sqrt{1-\rho^2} e^{h_1(Y_{t+1})} \varepsilon_{t+1}^2.$$

This is an instance of stochastic volatility models where $h_1(Y)$ corresponds to the log-volatility of an asset and X is the increment of the log-price of the same asset. Note that Y may have a slow autocorrelation decay (e.g. $a_j \sim j^{-\beta}$ with any $\beta > 1/2$ is possible). This model resembles the "fractional stochastic volatility model" of [4, 10]. Choose $x := 0$ and

$$F(w,s,z) := \kappa z + \rho e^{h_1(\sum_{j=0}^{\infty} a_j w_{t+1-j})} h_2(w_{t+1}) + \sqrt{1-\rho^2} e^{h_1(\sum_{j=0}^{\infty} a_j w_{t+1-j})} s.$$

As easily seen, Assumption 4.17 holds for this model and thus Theorem 4.19 below applies.

The functions h_1, h_2 serve as truncations only, in order to satisfy Assumption 4.17. One could probably relax Assumption 4.17 to accomodate the case $h_1(x) = h_2(x) = x$ as well. We refrain from the related complications in the present paper.

The result below permits to estimate the tracking error for another large class of non-Markovian processes. For simplicity, we consider only smooth functions H here.

Theorem 4.19. *Let $D \subset \mathbb{R}^N$ be bounded and open. Let Assumption 4.17 hold. Let $H : D \times \mathbb{R}^m \to \mathbb{R}^N$ be bounded, twice continuously differentiable, with bounded first and second derivatives. Then the conclusion of Theorem 3.6 is true for $H(\theta, X_t)$, $t \in \mathbb{N}$, $\theta \in D$, provided that Assumption 3.3 holds.*

The proof is given in Section 6. Most results in the literature are about homogeneous (controlled) Markov chains hence they do not apply to the present, inhomogeneous case and we exploit the L-mixing property in an essential way in our arguments. See, however, also Subsection 5.3 of [28] for alternative conditions in the inhomogeneous Markovian case.

5 Numerical implementation

Numerical results are presented here verifying the convergence properties of stochastic approximation procedures with a fixed gain in the case of discontinuous H, for Markovian and non-Markovian models. The purpose here is illustrative.

5.1 Quantile estimation for AR(1) processes

We first consider a Markovian example in the simplest possible case where $H(\theta, \cdot)$ is an indicator function. Let X_t, $t \in \mathbb{Z}$ be an AR(1) process defined by

$$X_{t+1} = \alpha X_t + \varepsilon_{t+1}$$

where α is a constant satisfying $|\alpha| < 1$ and ε_t, $t \in \mathbb{Z}$ are i.i.d standard normal variates. As a consequence of the above equation, one observes that

$$X_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

for every $t \in \mathbb{N}$. Moreover, X_t has stationary distribution which is $\nu := N(0, (1 - \alpha^2)^{-1})$ and the pair (X_t, X_{t+1}) has bivariate normal distribution with correlation α. We are interested in finding the quantile of the stationary distribution ν using the stochastic approximation method (1) with fixed gain.

The algorithm for the fixed gain $\lambda > 0$ is given by the following equation,

$$\theta_{t+1} = \theta_t + \lambda H(\theta_t, X_{t+1}), \tag{29}$$

for every $t \in \mathbb{N}$. For the purpose of the q-th quantile estimation of the stationary distribution ν, one takes

$$H(\theta, x) = q - 1_{\{x \leq \theta\}}. \tag{30}$$

With this choice of H, the solution of (13) is the quantile in question. The function H is just the gradient of the so-called "pinball" loss function introduced in Section 3 of [26] for quantile estimation. The true value of the q-th quantile of ν is $\Phi(q)/\sqrt{1-\alpha^2}$, where Φ is the cumulative distribution function of the standard normal variate. For our numerical experiments, we take $\alpha = 0.5$ and $q = 0.975$ and hence the true value of the q-th quantile is $\theta^* \approx 2.26$.

Figure 1 illustrates that the rate of convergence of the fixed gain algorithm is consistent with our theoretical findings in the paradigm of the quantile estimation of the stationary distribution of an AR(1) process. As noted above, the true value of the quantile in this particular example is 2.26 which is then compared with the estimate obtained by using the fixed gain approximation algorithm. The Monte Carlo estimate is based on 12000 samples and the number of iterations is taken to be $I = 10^6$ with initial value $\theta_0 = 2.0$.

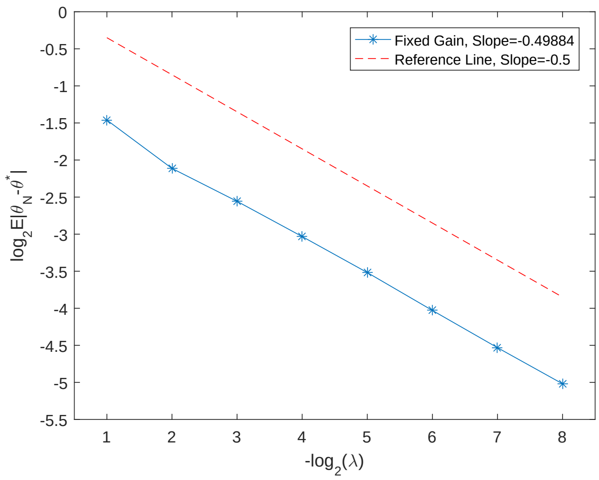

Figure 1: Rate of Convergence of Fixed Gain Algorithm for AR(1) Process

5.2 Quantile estimation for MA(∞) processes

Let us now consider the case when X_t, $t \in \mathbb{N}$ is an MA(∞) process which is non-Markovian. It is given by

$$X_t = \sum_{j=0}^{\infty} \frac{1}{(j+1)^{\beta}} \varepsilon_{t-j} \tag{31}$$

where $\beta > 1/2$ and ε_t, $t \in \mathbb{Z}$ are i.i.d sequence of standard normal variates. One can notice that the stationary distribution of MA(∞) process is given by

$$X_t \sim N\left(0, \sum_{j=0}^{\infty} \frac{1}{(j+1)^{2\beta}}\right)$$

for any $t \in \mathbb{N}$. As before, we are interested in the estimation of the quantile of the stationary distribution. In our numerical calculations, $\beta = 3$ and the exact variance is $\pi^6/945$. For generating the path of the MA(∞) process, we write X_t as

$$X_t = \sum_{j=0}^{t} \frac{1}{(j+1)^\beta} \varepsilon_{t-j} + \sum_{j=0}^{\infty} \frac{1}{(t+2+j)^\beta} \varepsilon_{-j-1}$$

and notice that

$$Y_t := \sum_{j=0}^{\infty} \frac{1}{(t+2+j)^\beta} \varepsilon_{-j-1} \sim N\Big(0, \sum_{j=0}^{\infty} \frac{1}{(t+2+j)^{2\beta}}\Big)$$

for any $t \in \mathbb{N}$. Also, a reasonable approximation of the variance of Y_t can be

$$\text{var}(Y_t) \approx \sum_{j=0}^{12} \frac{1}{(t+2+j)^{2\beta}}$$

which is within an interval of length 10^{-7} around the true value. With this set-up, the stochastic approximation method (29) with updating function (30) is implemented for the quantile estimation of the stationary distribution of MA(∞) process with $\theta_0 = 2.0$ and $\theta^* = 1.976950$ (0.975-th quantile). Figure 2 indicates that the rate of convergence of the fixed gain algorithm is 0.5, which is consistent with the theoretical findings. The Monte Carlo estimate is based on 10^5 samples. Figure 2 is based on $I = 10^5$ iterations.

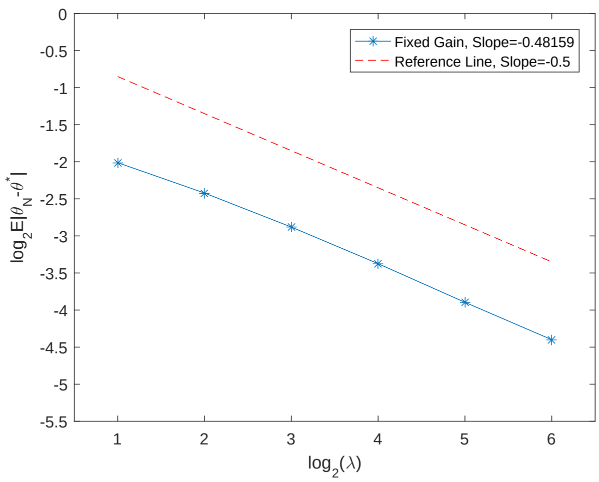

Figure 2: Rate of Convergence of Fixed Gain Algorithm for MA(∞) Process

5.3 Kohonen algorithm

In this section, we demonstrate the rate of convergence of the Kohonen algorithm for optimally quantizing a one-dimensional random variable X. We refer to [1, 9] for discussions. We fix the number of cells $N \geq 1$ in advance. Let $\theta := (\theta^1, \ldots, \theta^N) \in \mathbb{R}^N$ and define Voronoi cells as

$$\mathcal{V}^i(\theta) := \{x \in \mathbb{R} : |x - \theta^i| = \min_{j \in \{1,\ldots,N\}} |x - \theta^j|\}$$

for $i = 1, \ldots, N$. Values of X in a cell i will be quantized to θ^i. The zero-neighbourhood fixed gain Kohonen algorithm is aimed at minimizing, in θ, the quantity

$$\sum_{i+1}^{N} E\left[|X - \theta^i|^2 1_{\mathcal{V}^i(\theta)}(X)\right].$$

Differentiating (formally) this formula suggests the recursive procedure

$$\theta_{t+1}^i = \theta_t^i + \lambda \mathbf{1}_{\mathcal{V}^i(\theta)}(Y_t)(Y_t - \theta_t^i) \tag{32}$$

for every $i = 1, \ldots, N$ where $t \in \mathbb{N}$ and the process Y has a stationary distribution equal to the law of X. The algorithm approximates the $\mathbb{R}$-valued random variable X by θ^i if its values lie in the cell $\mathcal{V}^i(\theta)$, for every $i = 1, \ldots, N$.

In Figure 3, we demonstrate the rate of convergence of the zero-neighbourhood Kohonen algorithm with zero-neighbours when the signal Y_ts are i.i.d. observations from uniform distribution on $[0, 1]$, which is a well-understood case, see e.g. [1]. We take $N = 2$, $\theta := (\theta_1, \theta_2)$, $\mathcal{V}^1(\theta) = (0, (\theta^1 + \theta^2)/2]$ and $\mathcal{V}^2(\theta) = [(\theta_1 + \theta_2)/2, 1)$. Hence, the optimal value of θ is $\theta^{1*} = 1/4$ and $\theta^{2*} = 3/4$. The number of iterations is 10^8 and the number of sample paths is 10^3. Furthermore, the initial values of θs are $\theta_0^1 = 0.01$ and $\theta_0^2 = 0.02$. As illustrated, the rate of convergence is close to 0.5 which is consistent with the theoretical findings.

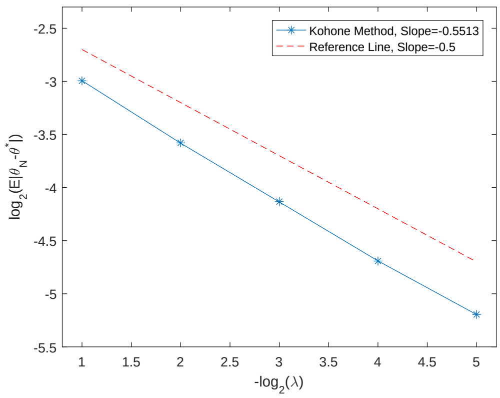

Figure 3: Rate of Convergence of Kohonen Algorithm for i.i.d. $U(0, 1)$.

Now, to have a non-Markovian example, consider a moving average process with lag 10, i.e.

$$X_t = \sum_{j=0}^{10} \frac{1}{(j+1)^\beta} \varepsilon_{t-j}, \ t \in \mathbb{N},$$

where ε_t, $t \in \mathbb{Z}$ are independent standard Gaussian random variables, denote it by MA(10). Clearly,

$$X_t \sim \mathcal{N}\Big(0, \sum_{j=0}^{10} \frac{1}{(j+1)^{2\beta}}\Big)$$

for any $t \geq 0$. Take $\beta := 3$ and notice that MA(10) is a good approximation of MA(∞) process (31) because the contributions from other terms are negligible due to low variance. We take $N = 2$ and implement the Kohonen algorithm (32) to sample two elements $\theta := (\theta^1, \theta^2)$ from the stationary distribution of the process Y defined by $Y_t := \tan^{-1}(X_t)$ for any $t \geq 0$. As the support of the stationary distribution of the process Y is $(-\pi/2, \pi/2)$, the Voronoi cells are $\mathcal{V}^1(\theta) := (-\pi/2, (\theta^1+\theta^2)/2]$ and $\mathcal{V}^2(\theta) := [(\theta^1 + \theta_2)/2, \pi/2)$. The true values $\theta^* := (\theta^{1*}, \theta^{2*})$ are the solution of the following system of two non-linear equations:

$$\theta^{1*}\Phi\Big(\frac{1}{\sigma}\tan\big(\frac{\theta^{1*}+\theta^{2*}}{2}\big)\Big) = E\Big(\tan^{-1}(\sigma Z)\mathbf{1}_{(-\infty,\frac{1}{\sigma}\tan(\frac{\theta^{1*}+\theta^{2*}}{2})]}(Z)\Big)$$

$$\theta^{2*}\Big[1-\Phi\Big(\frac{1}{\sigma}\tan\big(\frac{\theta^{1*}+\theta^{2*}}{2}\big)\Big)\Big] = E\Big(\tan^{-1}(\sigma Z)\mathbf{1}_{[\frac{1}{\sigma}\tan(\frac{\theta^{1*}+\theta^{2*}}{2}),\infty)}(Z)\Big)$$

where $\sigma^2 := \mathrm{var}(X_t)$, Z denotes the standard normal variate and Φ its distribution function.

Figure 4 is based on 10^8 iterations and 3000 paths (for Monte Carlo simulations). The initial values are $\theta_0^1 = -\pi/4$ and $\theta_0^2 = \pi/4$. Since θ^* is not known the output of the Kohonen algorithm (32) with $\lambda = 2^{-9}$ is taken as θ^*. Again, our numerical experiments are consistent with the theoretical rate $\lambda^{1/2}$ found in Theorem 3.6 above.

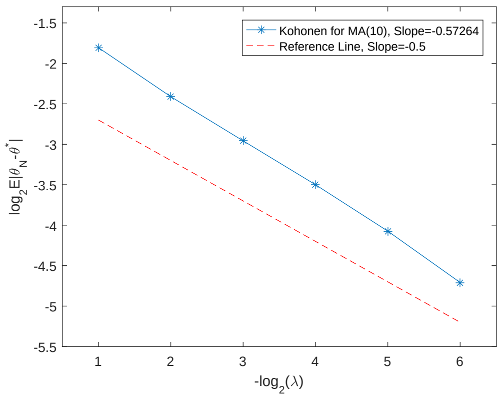

Figure 4: Rate of Convergence of Kohonen Algorithm for MA(10).

6 Appendix

Here we gather the proofs for Sections 2 and 3 as well as for Theorem 4.19. First we present a slight extension of Lemma 2.1 of [12] which is used multiple times.

Lemma 6.1. *Let $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ be sigma-algebras. Let X, Y be random variables in L^p such that Y is measurable with respect to $\mathcal{H} \vee \mathcal{G}$. Then for any $p \geq 1$,*

$$E^{1/p}\left[|X - E[X|\mathcal{H}\vee\mathcal{G}]|^p \middle| \mathcal{G}\right] \leq 2E^{1/p}\left[|X-Y|^p \middle| \mathcal{G}\right].$$

If Y is $\mathcal{H}$-measurable then

$$\|X - E[X|\mathcal{H}]\|_p \leq 2\|X - Y\|_p. \tag{33}$$

Proof. Since Y is $\mathcal{H}\vee\mathcal{G}$-measurable,

$$\begin{aligned}
E\left[|X-E\left[X|\mathcal{H}\vee\mathcal{G}\right]|^p\middle|\mathcal{G}\right] &\le \\
2^{p-1}E\left[|X-Y|^p\middle|\mathcal{G}\right]+2^{p-1}E\left[|Y-E\left[X|\mathcal{H}\vee\mathcal{G}\right]|^p\middle|\mathcal{G}\right] &\le \\
2^{p-1}E\left[|X-Y|^p\middle|\mathcal{G}\right]+2^{p-1}E\left[E\left[|Y-X|\middle|\mathcal{H}\vee\mathcal{G}\right]^p\middle|\mathcal{G}\right] &\le \\
2^{p-1}E\left[|X-Y|^p\middle|\mathcal{G}\right]+2^{p-1}E\left[E\left[|Y-X|^p\middle|\mathcal{H}\vee\mathcal{G}\right]\middle|\mathcal{G}\right] &\le \\
2^pE\left[|Y-X|^p\middle|\mathcal{G}\right],
\end{aligned}$$

by Jensen's inequality. Now (33) follows by taking $\mathcal{G}$ to be the trivial sigma-algebra. □

We now note what happens to products of two random fields.

Lemma 6.2. *Let $X_t(\theta)$ be ULM-rp and $Y_t(\theta)$ ULM-rq where $r\ge 1$, $p>1$, $1/p+1/q=1$. Then $X_t(\theta)Y_t(\theta)$ is ULM-r.*

Proof. We drop θ in the notation. It is clear from Hölder's inequality that

$$M_r(XY)\le M_{rp}(X)M_{rq}(Y),$$

so X_tY_t is bounded in L^r. Using Lemma 6.1, let us estimate, for $t,m\ge 1$,

$$\begin{aligned}
\|X_tY_t-E[X_tY_t|\mathcal{F}^+_{t-m}]\|_r &\le \\
2\|X_tY_t-E[X_t|\mathcal{F}^+_{t-m}]E[Y_t|\mathcal{F}^+_{t-m}]\|_r &\le \\
2\|X_tY_t-X_tE[Y_t|\mathcal{F}^+_{t-m}]\|_r+2\|X_tE[Y_t|\mathcal{F}^+_{t-m}]-E[X_t|\mathcal{F}^+_{t-m}]E[X_t|\mathcal{F}^+_{t-m}]\|_r &\le \\
2\|X_t\|_{rp}\|Y_t-E[Y_t|\mathcal{F}^+_{t-m}]\|_{rq}+2\|X_t-E[X_t|\mathcal{F}^+_{t-m}]\|_{rp}\|E[Y_t|\mathcal{F}^+_{t-m}]\|_{rq} &\le \\
2\|X_t\|_{rp}\|Y_t-E[Y_t|\mathcal{F}^+_{t-m}]\|_{rq}+2\|X_t-E[X_t|\mathcal{F}^+_{t-m}]\|_{rp}\|Y_t\|_{rq},
\end{aligned}$$

by Hölder's and Jensen's inequalities. This shows the L-mixing property of order r, noting the assumptions on X_t, Y_t. □

Lemma 6.3. *Let D be bounded. Fix $n\in\mathbb{N}$ and let ψ_t, $t\ge n$ be a sequence of D-valued, $\mathcal{F}_n$-measurable random variables. Let $X_t(\theta)$, $\theta\in D$, $t\in\mathbb{N}$ be UCLM-$(p,1)$ for some $p>1$, satisfying the CLC property and define the process $Y_t:=X_t(\psi_t)$, $t\ge n$. Then*

$$M^n_p(Y)\le M^n_p(X),\quad \Gamma^n_p(Y)\le\Gamma^n_p(X)\ a.s.$$

Proof. If the ψ_t are $\mathcal{F}_n$-measurable step functions then this follows easily from the definitions. For general ψ_t, one can take $\mathcal{F}_n$-measurable step function approximations ψ^k_t, $k\in\mathbb{N}$ of the ψ_t (in the almost sure sense). The CLC property implies that $X_t(\psi^k_t)$ tends to $X_t(\psi_t)$ in probability as $k\to\infty$. By Fatou's lemma, $M^n_p(Y(\psi^k_\cdot))\le M^n_p(X)$, $k\in\mathbb{N}$ now implies $M^n_p(Y(\psi_\cdot))\le M^n_p(X)$. The sequence $X_t(\psi^k_t)$ is bounded in L^p. It follows that $E[X_{n+t}(\psi^k_{n+t})|\mathcal{F}^+_{n+t-\tau}\vee\mathcal{F}_n]$ tends to $E[X_{n+t}(\psi_{n+t})|\mathcal{F}^+_{n+t-\tau}\vee\mathcal{F}_n]$ in L^1, a fortiori, in probability. Hence, for each $\tau\ge 1$, $\gamma^n_p(Y(\psi^k_\cdot),\tau)\le\gamma^n_p(X,\tau)$, $k\in\mathbb{N}$ implies $\gamma^n_p(Y(\psi_\cdot))\le\gamma^n_p(X,\tau)$, by Fatou's lemma. Consequently, $\Gamma^n_p(Y(\psi_\cdot))\le\Gamma^n_p(X)$ a.s. □

Remark 6.4. Fix $n\in\mathbb{N}$. Let Y_t be a conditionally L-mixing process of order $(p,1)$ for some $p\ge 1$ and define $W_t:=Y_t-E[Y_t|\mathcal{F}_n]$, $t\ge n$. Then it is easy to check that $M^n_p(W)\le 2M^n_p(Y)$ and $\Gamma^n_p(W)=\Gamma^n_p(Y)$.

Let us now enter the setting where for all $t \in \mathbb{N}$, $\mathcal{F}_t = \sigma(\varepsilon_j,\ j \in \mathbb{N},\ j \le t)$, $\mathcal{F}_t^+ := \sigma(\varepsilon_j,\ j > t)$ for some i.i.d. sequence ε_j, $j \in \mathbb{Z}$ with values in some Polish space $\mathcal{X}$. Let μ be the law of $(\varepsilon_0, \varepsilon_{-1}, \ldots)$ on $\mathcal{X}^{-\mathbb{N}}$. For given $\mathbf{e} = (e_0, e_{-1}, \ldots) \in \mathcal{X}^{-\mathbb{N}}$ and $n \in \mathbb{N}$, we define the measure

$$P^{\mathbf{e},n} := \left(\otimes_{i>n}\nu\right) \bigotimes \left(\otimes_{i\le n}\delta_{e_{i-n}}\right),$$

where δ_x is the probability concentrated on the point $x \in \mathcal{X}$. The corresponding expectation will be denoted by $E^{\mathbf{e},n}[\cdot]$.

In this setting the concept of conditional L-mixing is easily related to "ordinary" L-mixing and we will be able to use results of [12] directly, see the proof of Theorem 2.5. For each $n \in \mathbb{Z}$, we denote by Z_n the random variable $(\varepsilon_n, \varepsilon_{n-1}, \ldots)$ and by $\tilde{\mu}$ their law on $\mathcal{X}^{-\mathbb{N}}$ (which does not depend on n). Let X_t, $t \in \mathbb{N}$ be a *stochastic process* bounded in L^r for some $r \ge 1$. We introduce the quantities

$$\begin{aligned} M_r^{\mathbf{e},n}(X) &:= \sup_{t\in\mathbb{N}} E^{\mathbf{e},n}[|X_{n+t}|^r]^{1/r}, \\ \gamma_r^{\mathbf{e},n}(\tau,X) &:= \sup_{t\ge\tau} E^{\mathbf{e},n}[|X_{n+t} - E^{\mathbf{e},n}[X_{n+t}|\mathcal{F}^+_{n+t-\tau}]|^r]^{1/r},\ \tau \ge 1, \\ \Gamma_r^{\mathbf{e},n}(X) &:= \sum_{\tau=1}^{\infty} \gamma_r^{\mathbf{e},n}(\tau,X), \end{aligned}$$

which are well-defined for $\tilde{\mu}$-almost every $\mathbf{e}$.

Proof of Theorem 2.5. For any non-negative random variable Y on $(\Omega, \mathcal{F}, P)$,

$$E^{\mathbf{e},n}[Y]\Big|_{\mathbf{e}=Z_n} = E[Y|\mathcal{F}_n] \text{ a.s.} \tag{34}$$

This can easily be proved for indicators of the form $Y = 1_{\{\varepsilon_{n+j}\in A_j,\ -k\le j\le k\}}$ with some $k \in \mathbb{N}$ and with Borel sets $A_j \subset \mathcal{X}$ and then it extends to all non-negative measurable Y. It follows that

$$M_r^n(W) = M_r^{\mathbf{e},n}(W)|_{\mathbf{e}=Z_n}. \tag{35}$$

A similar argument also establishes

$$E^{\mathbf{e},n}[Y|\mathcal{F}^+_{n+t-\tau}]\Big|_{\mathbf{e}=Z_n} = E[Y|\mathcal{F}^+_{n+t-\tau} \vee \mathcal{F}_n] \text{ a.s.}, \tag{36}$$

for all $t \ge 1$ and $1 \le \tau \le t$ hence also

$$\gamma_r^{\mathbf{e},n}(\tau,X)|_{\mathbf{e}=Z_n} = \gamma_r^n(\tau,X) \text{ a.s.} \tag{37}$$

From the conditional L-mixing property of W_t, $t \in \mathbb{N}$ under P (of order $(r,1)$) it follows that, for $\tilde{\mu}$-almost every $\mathbf{e}$, the process W_{t+n}, $t \in \mathbb{N}$ is L-mixing under $P^{\mathbf{e},n}$. Theorems 1.1 and 5.1 of [12] (applied under $P^{\mathbf{e},n}$) imply

$$E^{\mathbf{e},n}\left[\max_{n<t\le m}\left|\sum_{s=n+1}^{t} b_s W_s\right|^r\right]^{1/r} \le C_r\left(\sum_{s=n+1}^{m} b_s^2\right)^{1/2}\sqrt{M_r^{\mathbf{e},n}(W)\Gamma_r^{\mathbf{e},n}(W)},$$

for $\tilde{\mu}$-almost every $\mathbf{e}$. Now (34), (35) and (37) imply (9). □

Now we turn to the proofs of Section 3. We first recall Lemma 2.2 of [17], which states that the discrete flow defined by (38) below inherits the exponential stability property (12). Let $\mathbb{M} := \{(m,n) \in \mathbb{N} : m \le n\}$.

Lemma 6.5. *Let Assumptions 3.2 and 3.3 be in force. For each $0 \le m \le n$ and $\xi \in D_\xi$, define $z : \mathbb{M} \times D \to D$ by the recursion*

$$z(m,m,\xi) := \xi, \quad z(n+1,m,\xi) := z(n,m,\xi) + \lambda G(z(n,m,\xi)). \tag{38}$$

If d is large enough and λ is small enough then this makes sense and $z(n,m,\xi) \in D_\theta$ for all $n \ge m$. Furthermore, for each $\alpha' < \alpha$ (see Assumption 3.3) there is $C(\alpha') > 0$ such that

$$\left| \frac{\partial}{\partial \xi} z(n,m,\xi) \right| \le C(\alpha') e^{-\lambda \alpha'(n-m)}. \tag{39}$$

□

Remark 6.6. Actually, the same arguments also imply that the recursion (38) is well-defined for all $\xi \in D_\theta$, stays in D and satisfies (39), provided that d' is large enough and λ is sufficiently small.

For convenience's sake, we recall a result from [11], which is also given as Lemma 4.2 of [17].

Lemma 6.7. *Let Assumptions 3.2 and 3.3 be satisfied. Let $y_t := y(t,0,\xi)$, $t \ge 0$. Let x_t, $t \ge 0$ be a continuous, piecewise continuously differentiable curve such that $x_0 = \xi$. Then for $t \ge 0$,*

$$x_t - y_t = \int_0^t \frac{\partial}{\partial \xi} y(t,w,x_w)(\dot{x}_w - G(x_w))dw. \tag{40}$$

Proof. For $0 \le w \le t$, let $z_w = y(t,w,x_w)$. The LHS of (40) can be written as

$$z_t - z_0 = \int_0^t \dot{z}_w dw = \int_0^t \left(\frac{\partial}{\partial w} y(t,w,x_w) + \frac{\partial}{\partial \xi} y(t,w,x_w)\dot{x}_w \right) dw.$$

From Theorem 3.1 on page 96 of [23] we obtain that, for all $x \in \mathbb{R}$,

$$\frac{\partial}{\partial w} y(t,w,x) + \frac{\partial}{\partial \xi} y(t,w,x)G(x) = 0,$$

and hence the proof is complete. □

Let $\xi \in D_\theta$ and define $\tilde{z}_n := z(n,0,\xi)$, $n \in \mathbb{N}$. The next lemma summarizes some arguments of [17] in the present setting, for the sake of a self-contained presentation.

Lemma 6.8. *Let Assumptions 3.2 and 3.3 be satisfied. Let $y_t := y(t,0,\xi)$ for some $\xi \in D_\xi$ and let θ_n be defined by* (15). *If d, d' are large enough then, for all $n \in \mathbb{N}$, we have $\theta_n \in D_\theta$ and also $\tilde{z}_n \in D$.*

Proof. We denote by θ_t the piecewise linear extension of θ_n, i.e. for $t \in (n, n+1)$, we set $\theta_t = (1-(t-n))\theta_n + (t-n)\theta_{n+1}$. For $w \in (n, n+1)$, it is easy to see that

$\dot{\theta}_w = \theta_{n+1} - \theta_n = \lambda H(\theta_{[w]}, X_{[w]+1})$ where $[w]$ denotes the integer part of w. Thus, Lemma 6.7 implies that as long as $\theta_w \in D_\theta$ for all $0 \le w \le t$,

$$\theta_t - y_t = \int_0^t \frac{\partial}{\partial \xi} y(t, w, \theta_w) \lambda \left(H(\theta_{[w]}, X_{[w]+1}) - G(\theta_w) \right) dw.$$

Since $|H|$ and $|G|$ are bounded by a constant, say, $C^\dagger$, (12) implies that

$$|\theta_t - y_t| \le \int_0^t C^* e^{-\lambda\alpha(t-w)} \lambda 2 C^\dagger dw \le 2 C^* C^\dagger \alpha^{-1}.$$

It is known that $y_t \in D_y$ whenever $y_0 \in D_\xi$. Now, if $d > 2C^* C^\dagger \alpha^{-1}$ then $|\theta_t - y_t|$ will be smaller than the distance between D_y and D_θ^c, where D_θ^c denotes the complement of D_θ, hence θ_t will stay in D_θ for ever.

The proof for $\tilde{z}_n \in D$ is similar. The piecewise linear extension of $\tilde{z}_n$ is denoted by $\tilde{z}_t$, $t \ge 0$. By computations as before,

$$\tilde{z}_t - y_t = \int_0^t \frac{\partial}{\partial \xi} y(t, w, \tilde{z}_w) \lambda \left(G(\tilde{z}_{[w]}) - G(\tilde{z}_w) \right) dw.$$

Denoting by K^* (resp. L^*) a bound for $|G|$ (resp. a Lipschitz-constant for G), we obtain

$$|G(\tilde{z}_{[w]}) - G(\tilde{z}_w)| \le L^* |\tilde{z}_{[w]} - \tilde{z}_w| \le L^* \lambda G(\tilde{z}_{[w]}) \le \lambda L^* K^*,$$

hence

$$|\tilde{z}_t - y_t| \le \int_0^t C^* e^{-\lambda\alpha(t-w)} \lambda^2 L^* K^* dw \le C^* \alpha^{-1} \lambda L^* K^*.$$

It follows that if $d' > C^* \alpha^{-1} \lambda L^* K^*$ then $\tilde{z}_t \in D$, for all t. □

Remark 6.9. Note that our estimates for d, d' in the above proof are somewhat different: by choosing λ small enough we can make d' as small as we wish whereas we do not have this option for d. This is in contrast with [17], where d can also be made arbitrarily small by choosing λ small. This difference comes from the fact that in [17] Lipschitz-continuity of $\theta \to H(\theta, \cdot)$ is assumed, unlike in the present setting.

Proof of Theorem 3.6. We follow the main lines of the arguments in [14, 17]. However, details deviate significantly as our present assumptions are different from those of the cited papers.

Lemma 6.8 above will guarantee that θ_t and $z_t, \overline{z}_t$ (see below) are well-defined. Clearly, $z_t = z(t, 0, \theta_0)$. Set $T = [1/(\lambda\alpha')]$, where $0 < \alpha' < \alpha$ is as in Lemma 6.5 and $[x]$ denotes the integer part of $x \in \mathbb{R}$. For each $n \in \mathbb{N}$, we set $\overline{z}_{nT} := \theta_{nT}$ and define recursively

$$\overline{z}_t := \overline{z}_{t-1} + \lambda G(\overline{z}_{t-1}), \qquad nT < t < (n+1)T.$$

In other words, $\overline{z}_t = z(t, nT, \theta_{nT})$. By the triangle inequality, we obtain, for any $t \in \mathbb{N}$,

$$|\theta_t - z_t| \le |\theta_t - \overline{z}_t| + |\overline{z}_t - z_t|. \tag{41}$$

Estimation for $|\theta_t - \overline{z}_t|$. Fix n and let $nT < t < (n+1)T$.

$$|\theta_t - \overline{z}_t| = \lambda \left| \sum_{k=nT}^{t-1} [H(\theta_k, X_{k+1}) - G(\overline{z}_k)] \right| \leq$$

$$\lambda \sum_{k=nT}^{t-1} \left| H(\theta_k, X_{k+1}) - H(\overline{z}_k, X_{k+1}) \right| +$$

$$\lambda \left| \sum_{k=nT}^{t-1} \left(H(\overline{z}_k, X_{k+1}) - E[H(\overline{z}_k, X_{k+1}) | \mathcal{F}_{nT}] \right) \right| +$$

$$\lambda \sum_{k=nT}^{t-1} \left| E[H(\overline{z}_k, X_{k+1}) | \mathcal{F}_{nT}] - G(\overline{z}_k) \right| =: \lambda(S_1 + S_2 + S_3).$$

It is clear that

$$ES_3 \leq E \left[\sup_{\vartheta \in D} \sum_{k=nT}^{\infty} \left| E[H(\vartheta, X_{k+1}) | \mathcal{F}_{nT}] - G(\vartheta) \right| \right] < C',$$

for some $C' < \infty$, by Assumption 3.4.

Turning our attention to S_1, the CLC property implies

$$ES_1 = E[E[S_1 | \mathcal{F}_{nT}]] \leq \sum_{k=nT}^{t-1} KE|\theta_k - \overline{z}_k|.$$

On each interval $nT \leq t < (n+1)T$, we now estimate S_2 as follows,

$$S_2 \leq \sup_{nT < t \leq (n+1)T} \left| \sum_{k=nT}^{t-1} \left(H(\overline{z}_k, X_{k+1}) - E[H(\overline{z}_k, X_{k+1}) | \mathcal{F}_{nT}] \right) \right|.$$

Note the UCLM-$(r,1)$ property of $H(\cdot,\cdot)$ as well as Lemma 6.3 and Remark 6.4. Apply Theorem 2.5 for nT instead of n and with the choice $b_t \equiv 1$ and

$$W_t := H(\overline{z}_t, X_{t+1}) - E[H(\overline{z}_t, X_{t+1}) | \mathcal{F}_{nT}], \ nT < t \leq (n+1)T, \ W_t := 0, \ 0 \leq t \leq nT,$$

note that $E[W_t | \mathcal{F}_{nT}] = 0$ for all t. We get

$$ES_2 = E[E[S_2 | \mathcal{F}_{nT}]] \leq E[E^{1/r}[S_2^r | \mathcal{F}_{nT}]] \leq$$

$$C_r T^{1/2} E\left[\sqrt{M_r^{nT}(W) \Gamma_r^{nT}(W)} \right] \leq C_r T^{1/2} \sqrt{EM_r^{nT}(W) E\Gamma_r^{nT}(W)} \leq$$

$$C'' T^{1/2}$$

with some $C'' < \infty$, independent of n, by the UCLM-$(r,1)$ property of W.

Putting together our estimates so far, we obtain for $nT \leq t < (n+1)T$,

$$E|\theta_t - \overline{z}_t| \leq \lambda \left(\sum_{k=nT}^{t-1} KE|\theta_k - \overline{z}_k| + C'' T^{1/2} + C' \right).$$

Recall that $E|\theta_t - \overline{z}_t|$ is finite by boundedness of D. The discrete Gronwall lemma yields the following estimate, independent of n:

$$E|\theta_t - \overline{z}_t| \leq \lambda (C'' T^{1/2} + C')(1 + \lambda K)^T. \tag{42}$$

Note that

$$(1+\lambda K)^T \le e^{\lambda KT} \le e^{K/\alpha'}.$$

Estimation for $|\bar{z}_t - z_t|$. Noting $z_0 = \theta_0$ and using the fundamental theorem of calculus, we estimate for $nT \le t < (n+1)T$, using telescoping sums,

$$\begin{aligned}
& |\bar{z}_t - z_t| \\
\le\ & \sum_{k=1}^{n} |z(t,kT,\theta_{kT}) - z(t,(k-1)T,\theta_{(k-1)T})| \\
=\ & \sum_{k=1}^{n} |z(t,kT,\theta_{kT}) - z(t,kT,z(kT,(k-1)T,\theta_{(k-1)T}))| \\
=\ & \sum_{k=1}^{n} \int_0^1 \left| \frac{\partial}{\partial \xi} z(t,kT,s\theta_{kT} + (1-s)z(kT,(k-1)T,\theta_{(k-1)T})) \right| ds \\
\times\ & |\theta_{kT} - z(kT,(k-1)T,\theta_{(k-1)T})| \\
\le\ & C(\alpha') \sum_{k=1}^{n} e^{-\lambda\alpha'(t-kT)} \left(|\theta_{kT-1} - \bar{z}_{kT-1}| + \lambda |H(\theta_{kT-1}, X_{kT}) - G(\bar{z}_{kT-1})| \right).
\end{aligned}$$

Notice that there is $\tilde{C} > 0$, independent of n, t such that

$$\sum_{k=1}^{n} e^{-\lambda\alpha'(t-kT)} \le \tilde{C}.$$

Therefore, the fact that H, G, D are bounded, imply

$$\begin{aligned}
E|\bar{z}_t - z_t| &\le c \sum_{k=1}^{n} e^{-\lambda\alpha'(t-kT)} E|\theta_{kT-1} - \bar{z}_{kT-1}| + c \sum_{k=1}^{n} e^{-\lambda\alpha'(t-kT)} \lambda \\
&\le c' \lambda^{1/2}, \qquad (43)
\end{aligned}$$

with some $c, c' > 0$, by (42) and by the choice of T. Finally, putting together our estimations (42), (43) and using (41), for λ small enough, we obtain

$$E|\theta_t - z_t| \le C\lambda^{1/2},$$

with some $C > 0$, which completes the proof. □

Proof of Corollary 3.7. Recall α' from Lemma 6.5. The fundamental theorem of calculus yields

$$\begin{aligned}
|z_t - \theta^*| &\le |z_0 - \theta^*| \int_0^1 \left| \frac{\partial}{\partial \xi} z(t,0,sz_0 + (1-s)\theta^*) \right| ds \\
&\le C(\alpha') e^{-\lambda\alpha' t} |z_0 - \theta^*|,
\end{aligned}$$

and this is $\le \lambda^{1/2}$ for $t \ge t_0(\lambda)$ if $t_0(\lambda) = C^\circ \ln(1/\lambda)/\lambda$ for some C°. Since

$$|\theta_t - \theta^*| \le |\theta_t - z_t| + |z_t - \theta^*|,$$

the statement follows. □

Proof of Theorem 4.19. Let us work conditionally on the event $\mathcal{E}_0 = \eta \in \mathbb{R}^{\mathbb{Z}}$ where

$$\mathcal{E}_l = (\varepsilon^1_{i+l})_{i\in\mathbb{Z}},$$

until further notice.

The CLC property and Assumption 3.2 are trivial. Define $\mathcal{F}_n := \sigma(\varepsilon^2_j; j \le n)$ and $\mathcal{F}^+_n := \sigma(\varepsilon^2_j; j > n)$.

We now prove that $H(\theta, X_t)$ is UCLM-$(r,1)$ with respect to the given $(\mathcal{F}_n, \mathcal{F}^+_n)$. Boundedness of H implies that $M^n_r(X)$, $n \in \mathbb{N}$ is uniformly bounded.

Fix $1 \le m \le t$. Define recursively

$$\xi_{t-m} := x, \quad \xi_{l+1} := F(\mathcal{E}_{l+1}, \varepsilon^2_{l+1}, \xi_l),\ l \ge t-m.$$

Set $X^+_{t,m} := \xi_t$. By construction, $X^+_{t,m}$ is $\mathcal{F}^+_{t-m}$-measurable and

$$|H(\theta, X^+_{t,m}) - H(\theta, X_t)| \le L\rho^m |x - X_{t-m}|,$$

where L is a Lipschitz-constant for $x \to H(\theta, x)$. So we can further estimate

$$\begin{aligned}
E\left[|x - X_{t-m}|^r \middle| \mathcal{F}_0\right]^{1/r} &\le \sum_{j=1}^{\infty} E\left[|\tilde{X}^{t-m}_j - \tilde{X}^{t-m}_{j-1}|^r \middle| \mathcal{F}_0\right]^{1/r} \le \\
\sum_{j=1}^{\infty} \rho^{j-1} E\left[|x - F(\mathbf{f}_{t-m-j+1}, \varepsilon^2_{t-m-j+1}, x)|^r \middle| \mathcal{F}_0\right]^{1/r} &\le \\
C\sum_{j=1}^{t-m} \rho^{j-1} \| |\varepsilon^2_{t-m-j+1}| + 1\|_r + C \sum_{j=t-m+1}^{\infty} \rho^{j-1}[|\varepsilon^2_{t-m-j+1}| + 1] &\le \\
C\| |\varepsilon^2_0| + 1\|_r \sum_{j=1}^{\infty} \rho^{j-1} + C\sum_{k=0}^{\infty} \rho^{t-m+k}[|\varepsilon^2_{-k}| + 1] &\le \\
C\| |\varepsilon^2_0| + 1\|_r \sum_{j=1}^{\infty} \rho^{j-1} + C\sum_{k=0}^{\infty} \rho^{k}[|\varepsilon^2_{-k}| + 1], &
\end{aligned}$$

using Assumption 4.17, the independence of ε^2_j, $j \ge 1$ from $\mathcal{F}_0$ and the $\mathcal{F}_0$-measurability of ε^2_j, $j \le 0$. Note that this last estimate is independent of t. We can carry out analogous estimates with $\mathcal{F}_n$ instead of $\mathcal{F}_0$ and these imply, via Lemma 6.1,

$$\gamma^n_r(m, X) \le 2LC\rho^m \left[(\|\varepsilon^2_0\|_r + 1) \sum_{j=1}^{\infty} \rho^{j-1} + \sum_{k=0}^{\infty} \rho^k [|\varepsilon^2_{n-k}| + 1] \right],$$

for each $n \in \mathbb{N}$, which implies that the sequence $\Gamma^n_r(X)$ is bounded in L^1, showing the UCLM-$(r,1)$ property for $H(\theta, X_t)$.

Since X^+_{t-m} is $\mathcal{F}^+_{t-m}$-measurable, the above estimates also show that $H(\theta, X_t)$ is (unconditionally) L-mixing of order $(r,1)$, hence Remark 3.5 implies Assumption 3.4. As the estimates are uniform in $\eta \in \mathbb{R}^{\mathbb{Z}}$, the argument of Theorem 3.6 can be applied. □

7 Conclusion

There is a large number of natural ramifications of our results that could be pursued: the estimation of higher order moments of the tracking error using the property UCLM-(r, p) for $p > 1$; accommodating multiple roots for equation (13); proving the convergence of the decreasing gain version of (1); considering the convergence of concrete procedures. We leave these for later work in order to convey a clear message, highlighting the novel techniques we have introduced.

Acknowledgments. We thank two anonymous referees for several insightful comments that led to substantial improvements. The major part of this work was done while the second author was working as a Whittaker Research Fellow in Stochastic Analysis in the School of Mathematics, University of Edinburgh, United Kingdom. We have made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF), see

`http://www.ecdf.ed.ac.uk/`

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1, in the framework of a "small research group", during the summer of 2016. Huy N. Chau and Miklós Rásonyi were also supported by the "Lendület" Grant LP2015-6 of the Hungarian Academy of Sciences and by the NKFIH (National Research, Development and Innovation Office, Hungary) grant KH 126505. Sotirios Sabanis gratefully acknowledges the support of the Royal Society through the IE150128 grant. We thank László Gerencsér for helpful discussions and dedicate this paper to him.

References

[1] M. Benaim, J.-C. Fort and G. Pagès. Convergence of the one-dimensional Kohonen algorithm. *Advances in Applied Probability*, 30:850-869, 1998.

[2] R. Bhattacharya and E. C. Waymire. An approach to the existence of unique invariant probabilities for Markov processes. In: *Limit theorems in probability and statistics, János Bolyai Math. Soc., I*, 181–200, 2002.

[3] J. A. Bucklew, T. G. Kurtz and W. A. Sethares. Weak convergence and local stability properties of fixed step size recursive algorithms. *IEEE Trans. Inform. Theory*, 39:966–978, 1993.

[4] F. Comte and É. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8:291–323, 1998.

[5] P. Diaconis and D. Freedman. Iterated random functions. *SIAM Review*, 41:45–76, 1999.

[6] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and convergence.* Wiley, New York, 1986.

[7] E. Eweda. Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data. *IEEE Transactions on Circuits and Systems*, 37: 1367-1374, 1990.

[8] E. Eweda. Convergence analysis of an adaptive filter equipped with the sign-sign algorithm. *IEEE Transactions on Automatic Control*, 40:1807-1811, 1995.

[9] G. Fort, É. Moulines, A. Schreck and M. Vihola. Convergence of Markovian stochastic approximation with discontinuous dynamics. *SIAM Journal on Control and Optimization*, 54:866–893, 2016.

[10] J. Gatheral, Th. Jaisson and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18:933–949, 2018.

[11] S. Geman. Some averaging and stability results for random differential equations. *SIAM Journal on Applied Mathematics*, 36:86–105, 1979.

[12] L. Gerencsér. On a class of mixing processes. *Stochastics*, 26:165–191, 1989.

[13] L. Gerencsér. AR(∞) estimation and nonparametric stochastic complexity. *IEEE Trans. Inform. Theory*, 38:1768–1778, 1992.

[14] L. Gerencsér. On fixed gain recursive estimation processes. *J. Mathematical Systems, Estimation and Control*, 6:355–358, 1996. Retrieval code for full electronic manuscript: 56854

[15] L. Gerencsér. Strong approximation of the recursive prediction error estimator of the parameters of an ARMA process. *Systems Control Lett.*, 21:347–351, 1993.

[16] L. Gerencsér. On Rissanen's predictive stochastic complexity for stationary ARMA processes. *J. Statist. Plann. Inference*, 41:303–325, 1994.

[17] L. Gerencsér. Stability of random iterative mappings. In: *Modeling uncertainty. An examination of stochastic theory, methods, and applications.* (ed. M. Dror, P. L'Écuyer, F. Szidarovszky), International Series in Operations Research and Management Science vol. 46, Kluwer Academic Publishers, 359–371, 2002.

[18] L. Gerencsér. A representation theorem for the error of recursive estimators. *SIAM J. Control Optim.*, 44:2123–2188, 2006.

[19] L. Gerencsér. Rate of convergence of recursive estimators. *SIAM J. Control Optim.*, 30:1200–1227, 1992.

[20] L. Gerencsér, G. Molnár-Sáska, Gy. Michaletzky, G. Tusnády and Zs. Vágó. New methods for the statistical analysis of Hidden Markov models. In:*Proceedings of the 41st IEEE Conference on Decision and Control, 2002, Las Vegas, USA* 2272–2277, IEEE Press, New York, 2002.

[21] L. Giraitis, H. L. Koul and D. Surgailis. *Large sample inference for long memory processes.* Imperial College Press, London, 2012.

[22] M. Hairer and J. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. *In: Seminar on stochastic analysis, random fields and applications VI (eds. R. Dalang, M. Dozzi and F. Russo)*, Progress in Probability, vol. 63, 109–117, 2011.

[23] Ph. Hartman. *Ordinary differential equations.* Classics in Applied Mathematics, 38, 2nd edition, SIAM, 2002.

[24] O. Kallenberg. *Foundations of modern probability.* 2nd edition, Springer, 2002.

[25] T. Kawata. *Fourier analysis in probability theory.* Academic Press, 1972.

[26] R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica*, 46:33-50, 1978.

[27] T. Kohonen. Analysis of a simple self-organising process. *Biological Cybernetics*, 44:135–140, 1982.

[28] S. Laruelle and G. Pagès. Stochastic approximation with averaging innovation applied to finance. *Monte Carlo Methods Appl.* 18:1–51, 2012.

[29] J.Neveu. *Discrete-parameter martingales.* North-Holland, 1975.

[30] M. Rásonyi. On the statistical analysis of quantized Gaussian AR(1) processes. *Int. J. of Adaptive Control and Signal Processing*, 24:490–507, 2010.