On the ergodicity of certain Markov chains in random environments*

Balázs Gerencsér[†] Miklós Rásonyi[‡]

November 26, 2018

Abstract

We study the ergodic behaviour of a discrete-time process X which is a Markov chain in a stationary random environment. The laws of X_t are shown to converge to a limiting law in (weighted) total variation distance as $t \to \infty$. Convergence speed is estimated and an ergodic theorem is established for functionals of X.

Our hypotheses on X combine the standard "small set" and "drift" conditions for geometrically ergodic Markov chains with conditions on the growth rate of a certain "maximal process" of the random environment. We are able to cover a wide range of models that have heretofore been untractable. In particular, our results are pertinent to difference equations modulated by a stationary Gaussian process. Such equations arise in applications, for example, in discretized stochastic volatility models of mathematical finance.

1 Introduction

Markov chains in random environments (*recursive chains* in the terminology of [4]) were systematically studied on countable state spaces in e.g. [5, 6, 20]. However, papers on the ergodic properties of such processes on a general state space are scarce and require rather strong, Doeblin-type conditions, see [16, 17, 21]. An exception is [22], where the system dynamics is assumed to be contracting instead but only weak convergence of the laws is established.

In this paper we deal with Markov chains in random environments that satisfy refinements of the usual hypotheses for the geometric ergodicity of Markov chains: minorization on "small sets", see Chapter 5 of [18], and Foster–Lyapunov type "drift" conditions, see Chapter 15 of [18].

Assuming that a suitably defined maximal process of the random environment satisfies a tail estimate, we manage to establish stochastic stability: ideas of [14] allow to obtain convergence to a limiting distribution in total variation norm with estimates on the convergence rate, see Sections 2 and 7 for the statements of our results. We also present a method to prove ergodic theorems, exploiting ideas of [1, 3, 13, 19], see Sections 2 and 7.

An important technical ingredient here is the notion of L-mixing, see Section 5. We present examples of difference equations modulated by Gaussian processes in Section 3. These can be regarded as discretizations of diffusions in random environments which arise, for instance, in stochastic volatility models of mathematical finance, see [7] and [10]. Proofs appear in Sections 4 and 6.

*Both authors enjoyed the support of the NKFIH (National Research, Development and Innovation Office, Hungary) grant KH 126505. The first author was also supported by the NKFIH grant PD 121107; the second author by the "Lendület" grant LP 2015-6 of the Hungarian Academy of Sciences and by The Alan Turing Institute, London under the EPSRC grant EP/N510129/1. We thank Attila Lovas for pointing out a mistake and for suggesting improvements. The paper also benefitted from comments by Nicolas Brosse, Éric Moulines, Sotirios Sabanis and Ramon van Handel.

†MTA Alfréd Rényi Institute of Mathematics and Eötvös Loránd University, Budapest, Hungary

‡MTA Alfréd Rényi Institute of Mathematics, Budapest, Hungary

2 Main results

Let $(\mathcal{Y}, \mathfrak{A})$ be a measurable space and let Y_t, $t \in \mathbb{Z}$ be a (strongly) stationary $\mathcal{Y}$-valued process on some probability space $(\Omega, \mathcal{F}, P)$. A generic element of Ω will be denoted by ω.

Expectation of a real-valued random variable X with respect to P will be denoted by $E[X]$ in the sequel. For $1 \leq p < \infty$ we write L^p to denote the Banach space of (a.s. equivalence classes of) $\mathbb{R}$-valued random variables with $E[|X|^p] < \infty$, equipped with the usual norm.

We fix another measurable space $(\mathcal{X}, \mathfrak{B})$ and denote by $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathfrak{B}$. Let $Q : \mathcal{Y} \times \mathcal{X} \times \mathfrak{B} \to [0,1]$ be a family of probabilistic kernels parametrized by $y \in \mathcal{Y}$, i.e. for all $A \in \mathfrak{B}$, $Q(\cdot,\cdot,A)$ is $\mathfrak{A} \otimes \mathfrak{B}$-measurable and for all $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $A \to Q(y,x,A)$ is a probability on $\mathfrak{B}$.

Let X_t, $t \in \mathbb{N}$ be a $\mathcal{X}$-valued stochastic process such that X_0 is independent of Y_t, $t \in \mathbb{Z}$ and

$$P(X_{t+1} \in A | \mathcal{F}_t) = Q(Y_t, X_t, A) \ P\text{-a.s.}, \ t \geq 0, \tag{1}$$

where the filtration is defined by

$$\mathcal{F}_t := \sigma(Y_j, \ j \leq t; \ X_j, \ 0 \leq j \leq t), \ t \geq 0.$$

Remark 2.1. Obviously, the law of X_t, $t \in \mathbb{N}$ (and also its joint law with Y_t, $t \in \mathbb{Z}$) are uniquely determined by (1). Let us consider the particular case where $\mathcal{X}$ is a Polish space with the corresponding family of Borel sets, $\mathfrak{B}$. Then, for every given Q, there exists a process X satisfying (1) (after possibly enlarging the probability space). See e.g. page 228 of [2] for a similar construction. We will establish a more precise result in Lemma 6.1 below, under additional assumptions.

The process Y will represent the random environment whose state Y_t at time t determines the transition law $Q(Y_t,\cdot,\cdot)$ of the process X at the given instant t. Our purpose is to study the ergodic properties of X.

We will now introduce a number of assumptions of various kinds that will figure in the statements of the main results: Theorems 2.11, 2.13, 2.14, 2.15, 7.1 and 7.2 below.

The following assumption closely resembles the well-known drift conditions for geometrically ergodic Markov chains, see e.g. Chapter 15 of [18]. In our case, however, there is also dependence on the state of the random environment.

Assumption 2.2. *(Drift condition) Let $V : \mathcal{X} \to [0,\infty)$ be a measurable function. Let $A_n \in \mathfrak{A}$, $n \in \mathbb{N}$ be a non-decreasing sequence of subsets such that $A_0 \neq \emptyset$ and $\mathcal{Y} = \cup_{n \in \mathbb{N}} A_n$. Define the $\mathbb{N}$-valued function*

$$\|y\| := \min\{n : \ y \in A_n\}, \ y \in \mathcal{Y}.$$

We assume that there is a non-increasing function $\lambda : \mathbb{N} \to (0,1]$ and a non-decreasing function $K : \mathbb{N} \to (0,\infty)$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\int_{\mathcal{X}} V(z)\, Q(y,x,dz) \leq (1 - \lambda(\|y\|))V(x) + K(\|y\|). \tag{2}$$

Furthermore, we may and will assume $\lambda(\cdot) \leq 1/3$ and $K(\cdot) \geq 1$.

We try to provide some intuition about Assumption 2.2: we expect that the stochastic process X behaves in an increasingly arbitrary way as the random environment Y becomes more and more "extreme" (i.e. $\|Y\|$ grows) so the drift condition (2) becomes less and less stringent on the increasing subsets A_n as n grows.

Example 2.3. A typical case is where $\mathcal{Y}$ is a subset of a Banach space $\mathbb{B}$ with norm $\|\cdot\|_{\mathbb{B}}$; $\mathfrak{A}$ its Borel field; $A_n := \{y \in \mathcal{Y} : \ \|y\|_{\mathbb{B}} \leq n\}$, $n \in \mathbb{N}$. In this setting

$$\|y\| = \lceil \|y\|_{\mathbb{B}} \rceil ,$$

where $\lceil\cdot\rceil$ stands for the ceiling function. In the examples of the present paper we will always have $\mathbb{B} = \mathbb{R}^d$ with some $d \geq 1$ and $|\cdot| = \|\cdot\|_{\mathbb{B}}$ will denote the respective Euclidean norm.

Another standard choice would be $\mathcal{Y} := \mathbb{N}$; $\mathfrak{A}$ is the power set of $\mathcal{Y}$; $A_n := \{i \in \mathbb{N} : \ i \leq n\}$. In this case $\|y\| = y$, $y \in \mathbb{N}$.

One more possibility could be $\mathcal{Y} := (0,\infty)$ with its Borel sets $\mathfrak{A}$ and with $A_n := [1/(n+1),\infty)$, $n \in \mathbb{N}$.

Remark 2.4. The reader will notice that we impose *nothing* about the ergodic behaviour of Y in our results, only estimates on its maximal process are required, see Assumption 2.7 below. It would be desirable to relax Assumption 2.2 allowing λ to vary in $(-\infty, 1)$ as long as "in the average" it is contractive (there are multiple options for the precise formulation of such a property). In that case, however, (strong) ergodic properties need to hold for Y. This is out of scope for the current work.

The next assumption stipulates the existence of a whole family of suitable "small sets" $C(R(n))$ that fit well the sets A_n appearing in Assumption 2.2.

Assumption 2.5. *(Minorization condition) For $R \geq 0$, set $C(R) := \{x \in \mathcal{X} : \; V(x) \leq R\}$. Let $\lambda(\cdot)$, $K(\cdot)$ be as in Assumption 2.2. Define $R(n) := 4K(n)/\lambda(n)$. There is a non-increasing function $\alpha : \mathbb{N} \to (0, 1]$ and for each $n \in \mathbb{N}$, there exists a probability measure ν_n on $\mathfrak{B}$ such that, for all $y \in \mathcal{Y}$, $x \in C(R(\|y\|))$ and $A \in \mathfrak{B}$,*

$$Q(y, x, A) \geq \alpha(\|y\|)\nu_{\|y\|}(A). \tag{3}$$

We may and will assume $\alpha(\cdot) \leq 1/3$.

In other words, depending on the "size" $\|y\|$ of state y of the random environment, we work on the set $C(4K(\|y\|)/\lambda(\|y\|))$ on which we are able to benefit from a "coupling effect" of strength $\alpha(\|y\|)$.

For a fixed V as in Assumption 2.2, let us define a family of metrics on

$$\mathcal{P}_V(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} V(x)\, \mu(dx) < \infty\}$$

by

$$\rho_\beta(\nu_1, \nu_2) := \int_{\mathcal{X}} [1 + \beta V(x)] |\nu_1 - \nu_2|(dx), \;\; \nu_1, \nu_2 \in \mathcal{P}_V(\mathcal{X}),$$

for each $0 \leq \beta \leq 1$. Here $|\nu_1 - \nu_2|$ is the total variation of the signed measure $\nu_1 - \nu_2$. Note that ρ_0 is just the total variation distance (and it can be defined for all $\nu_1, \nu_2 \in \mathcal{P}(\mathcal{X})$) while ρ_1 is the $(1 + V)$-weighted total variation distance.

Let $L : \mathcal{X} \times \mathfrak{B} \to [0, 1]$ be a probabilistic kernel. For each $\mu \in \mathcal{P}(\mathcal{X})$, we define the probability

$$[L\mu](A) := \int_{\mathcal{X}} L(x, A)\, \mu(dx), \;\; A \in \mathfrak{B}. \tag{4}$$

Consistently with these definitions, $Q(Y_n)\mu$ will refer to the action of the kernel $Q(Y_n, \cdot, \cdot)$ on μ. Note, however, that $Q(Y_n)\mu$ is a *random* probability measure.

For a bounded measurable function $\phi : \mathcal{X} \to \mathbb{R}$, we set

$$[L\phi](x) := \int_{\mathcal{X}} \phi(z)\, L(x, dz), \;\; x \in \mathcal{X}.$$

The latter definition makes sense for any non-negative measurable ϕ, too.

The following assumption is just an easily verifiable integrability condition about the initial values X_0 and X_1 of the process X.

Assumption 2.6. *(Second moment condition on the initial values)*

$$E\left[\left(\int_{\mathcal{X}} V(z)[\mu_0 + [Q(Y_0)\mu_0]](dz) \right)^2 \right] < \infty.$$

We now present a hypothesis controlling the maxima of $\|Y\|$ over finite time intervals (i.e. the "degree of extremity" of the random environment).

Assumption 2.7. *(Condition on the maximal process of the random environment) There exist a non-decreasing function $g : \mathbb{N} \to \mathbb{N}$ and a non-increasing function $\ell : \mathbb{N} \to [0, 1]$ such that*

$$P(\max_{1 \leq i \leq t} \|Y_i\| \geq g(t)) \leq \ell(t), \;\; t \geq 1. \tag{5}$$

Remark 2.8. It is clear that for a given process Y, several choices for the pair of functions g, ℓ are possible. Each of these leads to different estimates and it depends on Y and X which choice is better, no general rule can be determined a priori.

Remark 2.9. For Gaussian processes Y in $\mathcal{Y} := \mathbb{R}^d$, Assumption 2.7 holds, for instance, with $g(t) \sim \sqrt{t}$, $\ell(t) \sim \exp(-t)$, see Section 3 for more details.

Remark 2.10. One can derive estimates like (5) also for rather general processes Y. For instance, let Y_t, $t \in \mathbb{Z}$ be $\mathbb{R}^d$-valued strongly stationary such that $E|Y_0|^p < \infty$ for all $p \geq 1$. Then for each $q \geq 1$ set $p := 2q$ and estimate

$$\begin{aligned} & E^{1/q}\left[\max_{1\leq i\leq N}|Y_t|^q\right] \leq E^{1/2q}\left[\max_{1\leq i\leq N}|Y_t|^{2q}\right] \\ \leq\ & E^{1/2q}\left[\sum_{i=1}^N |Y_i|^{2q}\right] \leq C(q)N^{\frac{1}{2q}}, \end{aligned}$$

with constant $C(q) = E^{1/2q}[|Y_0|^{2q}]$. The Markov inequality implies that

$$P\left(\max_{1\leq i\leq N}|Y_t| \geq N\right) \leq \frac{C^q(q)N^{1/2}}{N^q} \leq \frac{C^q(q)}{N^{q-1/2}}. \tag{6}$$

Actually, for arbitrarily small $\chi > 0$ and arbitrarily large $r \geq 1$, we can set $q := \frac{r}{\chi} + \frac{1}{2}$ in (6) and then Assumption 2.7 holds with

$$g(k) := \lceil k^\chi \rceil \text{ and } \ell(k) := \frac{C^q(q)}{k^r},\ k \geq 1,$$

i.e. for arbitrary polynomially growing $g(\cdot)$ and polynomially decreasing $\ell(\cdot)$. This shows that our main results below have a wide spectrum of applicability well beyond the case of Gaussian Y, see Example 2.17.

We now define a number of quantities that will appear in various convergence rate estimates below. For each $t \in \mathbb{N}$, set

$$\begin{aligned} r_1(t) &:= \sum_{k=t}^\infty \frac{K(g(k))}{\alpha(g(k))} e^{-k\alpha(g(k))\lambda(g(k))/2}, \\ r_2(t) &:= \sum_{k=t}^\infty \frac{K(g(k+1))}{\alpha^2(g(k+1))\lambda(g(k+1))} \sqrt{\ell(k)}, \\ r_3(t) &:= \sum_{k=t}^\infty e^{-k\alpha(g(k))\lambda(g(k))/2}, \\ r_4(t) &:= \sum_{k=t}^\infty \ell(k), \\ \pi(t) &:= \frac{|\ln(\lambda(g(t)))|}{\alpha(g(t))\lambda(g(t))}. \end{aligned}$$

Introduce the notation $\mu_n := \mathrm{Law}(X_n)$, $n \in \mathbb{N}$. Now comes the first main result of the present paper: assuming our conditions on drift, minorization, initial values and control of the maxima, μ_t will tend to a limiting law as $t \to \infty$, provided that $r_1(0)$ and $r_2(0)$ are finite.

Theorem 2.11. *Let Assumptions 2.2, 2.5, 2.6 and 2.7 be in force. Assume*

$$r_1(0) + r_2(0) < \infty. \tag{7}$$

Then there is a probability μ_* on $\mathcal{X}$ such that $\mu_n \to \mu_*$ in $(1+V)$-weighted total variation as $n \to \infty$. More precisely,

$$\rho_1(\mu_n, \mu_*) \leq C[r_1(n) + r_2(n)],\ n \in \mathbb{N},$$

for some constant $C > 0$.

Theorem 2.13 below is just a variant of Theorem 2.11: under weaker assumptions it provides convergence in a weaker sense.

Assumption 2.12. *(First moment condition on the initial values)*

$$E\left[\int_{\mathcal{X}} V(z)[\mu_0 + [Q(Y_0)\mu_0]](dz)\right] < \infty.$$

Theorem 2.13. *Let Assumptions 2.2, 2.5, 2.7 and 2.12 be in force. Assume*

$$r_3(0) + r_4(0) < \infty. \tag{8}$$

Then there is a probability μ_* on $\mathcal{X}$ such that $\mu_n \to \mu_*$ in total variation as $n \to \infty$. More precisely,

$$\rho_0(\mu_n, \mu_*) \leq C[r_3(n) + r_4(n)],\ n \in \mathbb{N}, \tag{9}$$

for some constant $C > 0$.

Clearly, Assumption 2.6 implies Assumption 2.12 and (7) implies (8). Next, ergodic theorems corresponding to Theorems 2.11 and 2.13 are stated.

Theorem 2.14. *Let $\mathcal{X}$ be a Polish space and let $\mathfrak{B}$ be its Borel field. Let Assumptions 2.2, 2.5, 2.6 and 2.7 be in force, but with $R(n) := 8K(n)/\lambda(n)$, $n \in \mathbb{N}$ in Assumption 2.5. Let $\phi : \mathcal{X} \to \mathbb{R}$ be measurable such that*

$$|\phi(x)| \leq \tilde{C}(1 + V^{\delta}(x)),\ x \in \mathcal{X}, \tag{10}$$

for some $\tilde{C} > 0$ and $0 < \delta \leq 1/2$. Assume $r_1(0) + r_2(0) < \infty$ and

$$\left(\frac{K(g(N))}{\lambda(g(N))}\right)^{2\delta} \frac{\pi(N)}{N} \to 0,\ N \to \infty. \tag{11}$$

Then, for each $p < 1/\delta$,

$$\frac{\phi(X_1) + \ldots + \phi(X_N)}{N} \to \int_{\mathcal{X}} \phi(z)\mu_*(dz),\ N \to \infty \tag{12}$$

holds in L^p. (Here μ_* is the same as in Theorem 2.11 above.)

The rate of convergence in (12) can be estimated, see the proof of Theorem 2.14 in Section 6 below.

For bounded ϕ we have stronger results, under weaker assumptions.

Theorem 2.15. *Let $\mathcal{X}$ be a Polish space and let $\mathfrak{B}$ be its Borel field. Let Assumptions 2.2, 2.5, 2.7 and 2.12 be in force, but with $R(n) := 8K(n)/\lambda(n)$, $n \in \mathbb{N}$ in Assumption 2.5. Assume $r_3(0) + r_4(0) < \infty$. Let $\phi : \mathcal{X} \to \mathbb{R}$ be bounded and measurable. Then for every $p \geq 1$, L^p convergence in (12) holds whenever*

$$\pi(N)/N \to 0,\ N \to \infty. \tag{13}$$

Remark 2.16. In Theorems 2.14 and 2.15 above, we require a slight strenghtening of Assumption 2.5 by imposing (3) with $R(n) = 8K(n)/\lambda(n)$ instead of $R(n) = 4K(n)/\lambda(n)$.

Condition (13) is closely related to the condition $r_3(0) < \infty$ but none of the two implies the other. Indeed, fix $g(k) := k$. Choose λ constant and $\alpha(k) := \sqrt{\ln(k)k}$, $k \geq 4$. Then $\pi(k)/k \to 0$ but $r_3(0) = \infty$. Conversely, let $\alpha := 1/3$ and $\lambda(k) = \frac{12\ln(k)}{k}$. Then $r_3(0) < \infty$ but $\pi(k)/k$ tends to a positive constant.

Example 2.17. Let Y be strongly stationary $\mathbb{R}^d$-valued with $E|Y_0|^p < \infty$, $p \geq 1$. Let Assumptions 2.2 and 2.5 hold with $K(\cdot)$ having at most polynomial growth (i.e. $K(n) \leq cn^b$ with some $c, b > 0$) and $\alpha(\cdot)$, $\lambda(\cdot)$ having at most polynomial decay (i.e. $\alpha(n) \geq cn^{-b}$ with some $c, b > 0$, similarly for λ). Let Assumption 2.6 hold. Then Remark 2.10 shows (choosing χ small and r large) that Theorems 2.11 and 2.14 apply.

3 Examples about difference equations in Gaussian environments

In this section we present examples of processes X that satisfy a difference equation, modulated by the process Y. We do not aim at a high degree of generality but prefer to illustrate the power of the results in Section 2 in some easily tractable cases. We stress that, as far as we know, none of these results follow from the existing literature.

We fix $\mathcal{Y} := \mathbb{R}^d$ for some d and $\mathcal{X} := \mathbb{R}$. We also fix a $\mathcal{Y}$-valued zero-mean Gaussian stationary process Y_t, $t \in \mathbb{Z}$. We set $\|y\| := \lceil |y| \rceil$, $y \in \mathcal{Y}$ as in Example 2.3 above. We will exclusively use $V(x) := |x|$, $x \in \mathbb{R}$ in the examples below.

Remark 3.1. Let ξ_t, $t \in \mathbb{Z}$ be a zero-mean $\mathbb{R}$-valued stationary Gaussian process with unit variance. It is well-known that in this case

$$E\zeta_t \leq \sqrt{2\ln(t)} \leq \sqrt{2t},\ t \geq 1 \tag{14}$$

holds for $\zeta_t := \max_{1\leq i\leq t} \xi_i$. Furthermore, for all $a > 0$,

$$P(\zeta_t - E\zeta_t \geq a) \leq e^{-a^2/2}, \tag{15}$$

see [23, 24]. Applying (15) with $a := \sqrt{2t}$ and then proceeding analogously with the process $-\xi$, it follows from (14) that

$$P\left(\max_{1\leq i\leq t} |\xi_i| \geq 2\sqrt{2t}\right) \leq 2e^{-t}.$$

Applying these observations to every coordinate of Y, it follows that Assumption 2.7 holds for the process Y with the choice $g(k) := \lceil c_1\sqrt{k}\rceil$, $\ell(k) := \exp(-c_2 k)$ for some $c_1, c_2 > 0$ and thus $r_4(n)$ decreases at a geometric rate as $n \to \infty$.

More generally, choosing $a := t^b$ with some $b > 0$, Assumption 2.7 holds for Y with the choice $g(k) := \lceil c_1 k^b \rceil$, $\ell(k) := \exp(-c_2 k^{2b})$, by updating (14) and (15).

We assume throughout this section that ε_t, $t \in \mathbb{N}$ is an $\mathbb{R}$-valued i.i.d. sequence, independent of Y_t, $t \in \mathbb{Z}$; $E|\varepsilon_0|^2 < \infty$ and the law of ε_0 has an everywhere positive density f with respect to the Lebesgue measure, which is even and non-increasing on $[0, \infty)$. All these hypotheses could clearly be weakened/modified, we just try to stay as simple as possible.

Example 3.2. First we investigate the effect of the "contraction coefficient" λ in (2). Let $d := 1$. Let $0 < \underline{\sigma} \leq \overline{\sigma}$ be constants and $\sigma : \mathbb{R} \times \mathbb{R} \to [\underline{\sigma}, \overline{\sigma}]$ a measurable function. Let furthermore $\Delta : \mathbb{R} \to (0, 1]$ be even and non-increasing on $[0, \infty)$, for which we will develop conditions on the way. We stipulate that the tail of f is not too thin: it is at least as thick as that of a Gaussian variable, that is,

$$f(x) \geq e^{-sx^2},\ x \geq 0, \tag{16}$$

for some $s > 0$.

We assume that the dynamics of X is given by

$$X_0 := 0,\ X_{t+1} := (1 - \Delta(Y_t))X_t + \sigma(Y_t, X_t)\varepsilon_{t+1},\ t \in \mathbb{N}.$$

We will find $K(\cdot), \lambda(\cdot), \alpha(\cdot)$ such that Assumptions 2.2 and 2.5 hold and give an estimate for the rate $r_3(n)$ appearing in (9). (Note that we already have estimates for the rate $r_4(n)$ from Remark 3.1.)

The density of X_1 conditional to $X_0 = x$, $Y_0 = y$ (w.r.t. the Lebesgue measure) is easily seen to be

$$h_{x,y}(z) := f\left(\frac{z - (1 - \Delta(y))x}{\sigma(y, x)}\right)\frac{1}{\sigma(y, x)},\ z \in \mathbb{R}.$$

Fixing $\eta > 0$, we can estimate

$$\inf_{x,z\in[-\eta,\eta]} h_{x,y}(z) \geq f\left(\frac{2\eta}{\underline{\sigma}}\right)\frac{1}{\overline{\sigma}} =: m(\eta),$$

and $m(\cdot)$ does not depend on y. Define the probability measure

$$\nu_\eta(A) := \frac{1}{2\eta}\mathrm{Leb}(A \cap [-\eta,\eta]),\ A \in \mathfrak{B}.$$

It follows that

$$Q(y,x,A) \geq 2\eta m(\eta)\nu_\eta(A),\ A \in \mathfrak{B},$$

for all $x \in [-\eta,\eta]$, $y \in \mathbb{R}$. Notice that

$$[Q(y)V](x) \leq (1-\Delta(y))V(x) + \overline{\sigma}E|\varepsilon_0| \leq (1-\Delta(y))V(x) + K,$$

where $K := \max\{\overline{\sigma}E|\varepsilon_0|, 1\}$. Then Assumption 2.2 holds with $A_n := \{x \in \mathbb{R} : |x| \leq n\}$, $\lambda(n) := \Delta(n)$ and $K(n) := K$, $n \geq 1$. (Here and in the sequel we use the index set $\mathbb{N} \setminus \{0\}$ instead of $\mathbb{N}$ for convenience.)

Let $\eta := \tilde{R}(y) := 4K/\Delta(y)$, $y \in \mathcal{Y}$ and $R(n) := \tilde{R}(n)$, $n \in \mathbb{N}$. We note that $\tilde{R}(y)$ is defined for every $y \in \mathcal{Y}$ while $R(n)$ is defined for every $n \in \mathbb{N}$, this is why we keep different notations for these two functions here and also in the subsequent examples. We can conclude, using the tail bound (16) that

$$Q(y,x,A) \geq \frac{8Km(\tilde{R}(y))}{\Delta(y)}\nu_{\tilde{R}(y)}(A) \geq \frac{e^{-c_3\tilde{R}^2(y)}}{\Delta(y)}\nu_{\tilde{R}(y)}(A),$$

for all $A \in \mathfrak{B}$, with some $c_3 > 0$ so (3) in Assumption 2.5 holds with

$$\alpha(n) := e^{-c_3R^2(n)}/\Delta(0),\ n \geq 1,$$

and $\nu_n := \nu_{R(n)}$. Now let the function Δ be such that $\Delta(y) := 1$ for $0 \leq y < 3$ and $\Delta(y) \geq 1/(\ln(y))^\delta$ with some $\delta > 0$, for all $y \geq 3$. We obtain from the previous estimates and from Remark 3.1 with $g(k) = \lceil c_1\sqrt{k}\rceil$ that

$$\lambda(g(k))\alpha(g(k)) \geq e^{-c_4\ln^{2\delta}(k)},$$

with some $c_4 > 0$. When $\delta < 1/2$, this leads to estimates on the terms of $r_3(n)$ which guarantee $r_3(0) < \infty$.

If instead of (16) we assume

$$f(x) \geq e^{-sx},\ x \geq 0,$$

then $r_3(0) < \infty$ follows whenever $\delta < 1$. This shows nicely the interplay between the feasible fatness of the tail of f and the strength of the mean-reversion $\Delta(\cdot)$.

Example 3.3. Again, let $d := 1$, $X_0 := 0$ and

$$X_{t+1} := (1-\Delta)X_t + \sigma(Y_t, X_t)\varepsilon_{t+1},\ t \in \mathbb{N},$$

where $\sigma : \mathbb{R} \times \mathbb{R} \to (0,\infty)$ is a measurable function and $0 \leq \Delta < 1$ is a constant. We furthermore assume that

$$c_5G(y) \leq \sigma(y,x) \leq c_6G(y),\ x \in \mathbb{R},$$

with some even function $G : \mathbb{R} \to (0,\infty)$ that is nondecreasing on $[0,\infty)$ and with constants $c_5, c_6 > 0$. We clearly have (2) with $\lambda(n) := \Delta$, $n \in \mathbb{N}$ (i.e. $\lambda(\cdot)$ is constant) and $A_n := \{x \in \mathbb{R} : |x| \leq n\}$, $K(n) := \tilde{K}(n)$, $n \in \mathbb{N}$ where $\tilde{K}(y) := c_6G(y)E|\varepsilon_0|$, $y \in \mathbb{R}$. Taking $\tilde{R}(y) := 4\tilde{K}(y)/\Delta$, $y \in \mathbb{R}$, estimates as in Example 3.2 lead to

$$Q(y,x,A) \geq 2\tilde{R}(y)f\left(\frac{2\tilde{R}(y)}{c_5G(y)}\right)\frac{1}{c_6G(y)}\nu_{\tilde{R}(y)}(A) \geq c_7\nu_{\tilde{R}(y)}(A),$$

for all $A \in \mathfrak{B}$ with some fixed constant $c_7 > 0$, where $\nu_{\tilde{R}(y)}(\cdot)$ is the normalized Lebesgue measure restricted to $C(\tilde{R}(y))$, as in Example 3.2 above, so setting $R(n) := \tilde{R}(n)$, $n \in \mathbb{N}$, we can choose $\nu_n := \nu_{R(n)}$ and $\alpha(\cdot)$ a positive constant.

Assume e.g., $G(y) \leq C[1 + |y|^q]$, $y \geq 0$ with some $C, q > 0$ and choose $g(k) := \lceil c_1\sqrt{k}\rceil$, $\ell(k) := \exp(-c_2k)$, as discussed in Remark 3.1. Then Theorems 2.11 and 2.14 apply.

Example 3.4. We now investigate a discrete-time model for financial time series, inspired by the "fractional stochastic volatility model" of [7, 10].

Let w_t, $t \in \mathbb{Z}$ and ε_t, $t \in \mathbb{N}$ be two sequences of i.i.d. random variables such that the two sequences are also independent. Assume that w_t are Gaussian. We define the (causal) infinite moving average process

$$\xi_t := \sum_{j=0}^{\infty} a_j w_{t-j},\ t \in \mathbb{Z}.$$

This series is almost surely convergent whenever $\sum_{j=0}^{\infty} a_j^2 < \infty$. We take $d := 2$ here and the random environment will be the $\mathcal{Y} = \mathbb{R}^2$-valued process $Y_t := (w_t, \xi_t)$, $t \in \mathbb{Z}$.

We imagine that ξ_t describes the *log-volatility* of an asset in a financial market. It is reasonable to assume that ξ is a Gaussian linear process (see [10] where the related continuous-time models are discussed in detail).

Let us now consider the $\mathbb{R}$-valued process X which will describe the *increment of the log-price* of the given asset. Assume that $X_0 := 0$,

$$X_{t+1} = (1-\Delta)X_t + \rho e^{\xi_t} w_t + \sqrt{1-\rho^2} e^{\xi_t} \varepsilon_{t+1},\ t \in \mathbb{N},$$

with some $-1 < \rho < 1$, $0 < \Delta \leq 1$. The logprice is thus jointly driven by the noise sequences ε_t, w_t. The parameter Δ is responsible for the autocorrelation of X (Δ is typically close to 1). The parameter ρ controls the correlation of the price and its volatility. This is found to be non-zero (actually, negative) in empirical studies, see [8], hence it is important to include w_t, $t \in \mathbb{Z}$ both in the dynamics of X and in that of Y. We take $A_n := \{y = (w, \xi) \in \mathbb{R}^2 :\ |y| \leq n\}$, $n \in \mathbb{N}$.

Notice that

$$|X_1| \leq (1-\Delta)|X_0| + [|w_0| + |\varepsilon_1|]e^{\xi_t}$$

hence

$$E[V(X_1)|X_0 = x,\ Y_0 = (w, \xi)] \leq (1-\Delta)V(x) + c_8 e^{\xi}(1 + |w|)$$

for all $x \in \mathbb{R}$, with some $c_8 > 0$, i.e. Assumption 2.2 holds with $\lambda(n) := \lambda := \Delta$ and $K(n) := c_8 e^n (1+n)$.

We now turn our attention to Assumption 2.5. Denote the density of the law of X_1 conditional to $X_0 = x$, $Y_0 = (w, \xi)$ with respect to the Lebesgue measure by $h_{x,w,\xi}(z)$, $z \in \mathbb{R}$. For $x, z \in [-\eta, \eta]$ we clearly have

$$h_{x,w,\xi}(z) \geq f\left(\frac{2\eta + e^{\xi}|w|}{e^{\xi}\sqrt{1-\rho^2}}\right) \frac{1}{e^{\xi}\sqrt{1-\rho^2}}. \tag{17}$$

We assume from now on that f, the density of ε_0 satisfies

$$f(x) \geq s/(1+x)^{\chi},\ x \geq 0$$

with some $s > 0$, $\chi > 3$, this is reasonable as X_t has fat tails according to empirical studies, see [8]. At the same time, Assumption 2.6 can also be satisfied for such a choice of f.

Define $\tilde{K}(y) := e^{\xi}(1 + |w|)$ and $\tilde{R}(y) := 4\tilde{K}(y)/\lambda$, for $y = (w, \xi) \in \mathbb{R}^2$. Use (17) to obtain, as in Example 3.2 above,

$$Q(y, x, A) \geq \frac{c_9}{(1+|w|)^{\chi}} \frac{1}{e^{\xi}} 2\tilde{R}(y) \nu_{\tilde{R}(y)}(A) \geq \frac{c_{10}}{(1+|w|)^{\chi-1}} \nu_{\tilde{R}(y)}(A),$$

with fixed constants $c_9, c_{10} > 0$, where ν_η is the normalized Lebesgue measure restricted to $[-\eta, \eta]$. Set $R(n) := \tilde{R}((n, n))$, $n \geq 1$. Then Assumption 2.5 holds with

$$\alpha(n) := \frac{c_{10}}{(1+n)^{\chi-1}},\ n \geq 1.$$

Recalling the end of Remark 3.1, and choosing $b > 0$ small enough we can conclude that Theorems 2.11 and 2.14 apply to this stochastic volatility model.

Although the examples above are rather elementary and restricted in their scope, they point towards large classes of models, relevant in applications, where the results of Section 2 apply in a powerful way.

4 Proofs of stochastic stability

We first present a result of [14] (see also the related ideas in [15]) which will be used below.

Lemma 4.1. *Let $L : \mathcal{X} \times \mathfrak{B} \to [0, 1]$ be a probabilistic kernel such that*

$$LV(x) \le \gamma V(x) + K, \; x \in \mathcal{X},$$

for some $0 \le \gamma < 1$, $K > 0$. Let $C := \{x \in \mathcal{X} : V(x) \le R\}$ for some $R > 2K/(1-\gamma)$. Let us assume that there is a probability ν on $\mathfrak{B}$ such that

$$\inf_{x \in C} L(x, A) \ge \alpha \nu(A), \; A \in \mathfrak{B},$$

for some $\alpha > 0$. Then for each $\alpha_0 \in (0, \alpha)$ and for $\gamma_0 := \gamma + 2K/R$,

$$\rho_\beta(L\mu_1, L\mu_2) \le \max\left\{1 - (\alpha - \alpha_0), \frac{2 + R\beta\gamma_0}{2 + R\beta}\right\} \rho_\beta(\mu_1, \mu_2), \; \mu_1, \mu_2 \in \mathcal{P}_V,$$

holds for $\beta = \alpha_0/K$. □

For the proof, see Theorem 3.1 in [14]. Next comes an easy corollary.

Lemma 4.2. *Let $L : \mathcal{X} \times \mathfrak{B} \to [0, 1]$ be a probabilistic kernel such that*

$$LV(x) \le (1 - \lambda)V(x) + K, \; x \in \mathcal{X}, \tag{18}$$

for some $0 < \lambda \le 1/3$, $K > 0$. Let $C := \{x \in \mathcal{X} : V(x) \le R\}$ with $R := 4K/\lambda$. Assume that there is a probability ν on $\mathfrak{B}$ such that

$$\inf_{x \in C} L(x, A) \ge \alpha \nu(A), \; A \in \mathfrak{B}, \tag{19}$$

for some $0 < \alpha \le 1/3$. Then

$$\rho_\beta(L\mu_1, L\mu_2) \le \left(1 - \frac{\alpha\lambda}{2}\right) \rho_\beta(\mu_1, \mu_2), \; \mu_1, \mu_2 \in \mathcal{P}_V,$$

holds for $\beta = \alpha/2K$.

Proof. Choose $\gamma := 1-\lambda$, and let $\alpha_0 := \alpha/2$. Note that $1-(\alpha-\alpha_0) = 1-\alpha/2$ and $R\beta = 4\alpha_0/(1-\gamma)$ holds for $\beta = \alpha_0/K$. Applying Lemma 4.1, we estimate

$$\begin{aligned} \rho_\beta(L\mu_1, L\mu_2) &\le \\ \max\left\{1 - (\alpha - \alpha_0), \frac{2 + R\beta\gamma_0}{2 + R\beta}\right\} \rho_\beta(\mu_1, \mu_2) &= \\ \max\left\{1 - \alpha/2, 1 - \frac{4\alpha_0(1-\gamma_0)/(1-\gamma)}{2 + 4\alpha_0/(1-\gamma)}\right\} \rho_\beta(\mu_1, \mu_2). \end{aligned}$$

Here

$$\frac{4\alpha_0(1-\gamma_0)/(1-\gamma)}{2 + 4\alpha_0/(1-\gamma)} = \frac{\alpha_0\lambda}{\lambda + 2\alpha_0} \ge \alpha_0\lambda$$

and we get the statement since $\alpha/2 \ge \alpha_0\lambda$. □

Let $(\mathcal{T}, \mathfrak{T})$ be some measurable space. When $(x, A) \to L(x, A)$, $x \in \mathcal{T}$, $A \in \mathfrak{B}$ is a (not necessarily probabilistic) kernel and Z is a $\mathcal{T}$-valued random variable then we define a measure $\mathcal{E}[L(Z)](\cdot)$ on $\mathcal{B}$ via

$$\mathcal{E}[L(Z)](A) := E[L(Z, A)], \; A \in \mathfrak{B}. \tag{20}$$

We will use the following trivial inequalities in the sequel:

$$\rho_0(\cdot) \le 2, \quad \rho_0(\cdot) \le \rho_\beta(\cdot) \le \rho_1(\cdot) \le \left(1 + \frac{1}{\beta}\right) \rho_\beta(\cdot), \; 0 < \beta \le 1. \tag{21}$$

Proof of Theorem 2.11. Fix $\mathbf{y} := (y_0, y_{-1}, y_{-2}, \ldots) \in \mathcal{Y}^{-\mathbb{N}}$ for the moment. Let $\mathbf{y}_n := (y_0, y_{-1}, \ldots, y_{-n+1})$, $n \geq 1$, set

$$\mu_n(\mathbf{y}_n) := Q(y_0)Q(y_{-1}) \ldots Q(y_{-n+1})\mu_0, \ n \geq 1.$$

Here $Q(y)$ is the operator acting on probabilities which is described in (4) above but, instead of $L(x, A)$, with the kernel $Q(y, x, A)$.

Fix $n \geq 1$ and denote $\bar{y}_n := \max_{-n+1 \leq j \leq 0} \|y_j\|$. Since

$$\alpha(\|y_j\|) \geq \alpha(\bar{y}_n), \ \lambda(\|y_j\|) \geq \lambda(\bar{y}_n), \ K(\|y_j\|) \leq K(\bar{y}_n),$$

for each $-n+1 \leq j \leq 0$, (18) and (19) hold for $L = Q(y_j)$, $j = -n+1, \ldots, 0$ with $K = K(\bar{y}_n)$, $\lambda = \lambda(\bar{y}_n)$ and $\alpha = \alpha(\bar{y}_n)$. An n-fold application of Lemma 4.2 implies that, for $\beta = \alpha(\bar{y}_n)/2K(\bar{y}_n)$,

$$\rho_\beta(\mu_n(\mathbf{y}_n), \mu_{n+1}(\mathbf{y}_{n+1})) \leq (1 - \alpha(\bar{y}_n)\lambda(\bar{y}_n)/2)^n \rho_\beta(\mu_0, Q(y_{-n})\mu_0).$$

By (21) and by $K(\cdot)/\alpha(\cdot) \geq 1$,

$$\begin{aligned} \rho_1(\mu_n(\mathbf{y}_n), \mu_{n+1}(\mathbf{y}_{n+1})) &\leq \\ \left(1 + \frac{2K(\bar{y}_n)}{\alpha(\bar{y}_n)}\right)(1 - \alpha(\bar{y}_n)\lambda(\bar{y}_n)/2)^n \rho_\beta(\mu_0, Q(y_{-n})\mu_0) &\leq \\ \frac{3K(\bar{y}_n)}{\alpha(\bar{y}_n)}(1 - \alpha(\bar{y}_n)\lambda(\bar{y}_n)/2)^n \rho_1(\mu_0, Q(y_{-n})\mu_0).& \end{aligned}$$

Now let $\mathbf{Y}_n := (Y_0, Y_{-1}, \ldots, Y_{-n+1})$. In the sequel we will need the definition (20) for the kernel $(\mathbf{z}, A) \to \mu_n(\mathbf{z})(A)$, $\mathbf{z} \in \mathcal{Y}^n$, $A \in \mathfrak{B}$ (and for similar kernels). Notice that, for any measurable function $w : \mathcal{X} \to \mathbb{R}_+$,

$$\int_{\mathcal{X}} w(z) \ |\mathcal{E}[\mu_n(\mathbf{Y}_n)] - \mathcal{E}[\mu_{n+1}(\mathbf{Y}_{n+1})]| \, (dz) \leq \int_{\mathcal{X}} w(z) \, \mathcal{E}\left[|\mu_n(\mathbf{Y}_n) - \mu_{n+1}(\mathbf{Y}_{n+1})|\right](dz).$$

This is trivial for indicators and then follows for all measurable w in a standard way. By similar arguments, we also have

$$\int_{\mathcal{X}} w(z)\mathcal{E}\left[|\mu_n(\mathbf{Y}_n) - \mu_{n+1}(\mathbf{Y}_{n+1})|\right](dz) = E\left[\int_{\mathcal{X}} w(z) \, |\mu_n(\mathbf{Y}_n) - \mu_{n+1}(\mathbf{Y}_{n+1})| \, (dz)\right].$$

Since $\mu_n = \mathcal{E}[\mu_n(\mathbf{Y}_n)]$, we infer that

$$\begin{aligned} \rho_1(\mu_n, \mu_{n+1}) = \int_{\mathcal{X}} (1 + V(z)) \, |\mathcal{E}[\mu_n(\mathbf{Y}_n)] - \mathcal{E}[\mu_{n+1}(\mathbf{Y}_{n+1})]| \, (dz) &\leq \\ \int_{\mathcal{X}} (1 + V(z))\mathcal{E}\left[|\mu_n(\mathbf{Y}_n) - \mu_{n+1}(\mathbf{Y}_{n+1})|\right](dz) &= \\ E\left[\int_{\mathcal{X}} (1 + V(z)) \, |\mu_n(\mathbf{Y}_n) - \mu_{n+1}(\mathbf{Y}_{n+1})| \, (dz)\right] &= \\ E[\rho_1(\mu_n(\mathbf{Y}_n), \mu_{n+1}(\mathbf{Y}_{n+1}))].& \end{aligned}$$

We thus arrive at

$$\rho_1(\mu_n, \mu_{n+1}) \leq 3E\left[\frac{K(M_n)}{\alpha(M_n)}(1 - \alpha(M_n)\lambda(M_n)/2)^n \rho_1(\mu_0, Q(Y_{-n})\mu_0)\right], \tag{22}$$

using the notation $M_n := \max_{-n+1 \leq i \leq 0} \|Y_i\|$.

We now estimate the expectation on the right-hand side of (22) separately on the events $\{M_n \geq g(n)\}$ and $\{M_n < g(n)\}$.

Note that

$$\begin{aligned} & E\left[\frac{K(M_n)}{\alpha(M_n)}\left(1 - \frac{\alpha(M_n)\lambda(M_n)}{2}\right)^n \rho_1(\mu_0, Q(Y_{-n})\mu_0)1_{\{|M_n| \geq g(n)\}}\right] \\ \leq & \sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))}\left(1 - \frac{\alpha(g(k+1))\lambda(g(k+1))}{2}\right)^n E\left[\rho_1(\mu_0, Q(Y_{-n})\mu_0)1_{\{g(k+1) > |M_n| \geq g(k)\}}\right] \\ \leq & \sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))}\left(1 - \frac{\alpha(g(k+1))\lambda(g(k+1))}{2}\right)^n E\left[\rho_1(\mu_0, Q(Y_{-n})\mu_0)1_{\{|M_n| \geq g(k)\}}\right]. \end{aligned}$$

Hence

$$
\begin{aligned}
&\sum_{m=n}^{\infty} \rho_1(\mu_m, \mu_{m+1}) \\
\leq\ & 3\sum_{m=n}^{\infty} \frac{K(g(m))}{\alpha(g(m))} e^{-\frac{m}{2}\alpha(g(m))\lambda(g(m))} E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0) 1_{\{|M_m|<g(m)\}}\right] \\
+\ & 3\sum_{m=n}^{\infty}\sum_{k=m}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))} \left(1-\frac{\alpha(g(k+1))\lambda(g(k+1))}{2}\right)^m E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0) 1_{\{|M_m|\geq g(k)\}}\right] \\
\leq\ & 3\sum_{m=n}^{\infty} \frac{K(g(m))}{\alpha(g(m))} e^{-\frac{m}{2}\alpha(g(m))\lambda(g(m))} E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0)\right] \\
+\ & 3\sum_{k=n}^{\infty}\sum_{m=n}^{k} \frac{K(g(k+1))}{\alpha(g(k+1))} \left(1-\frac{\alpha(g(k+1))\lambda(g(k+1))}{2}\right)^m E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0) 1_{\{|M_m|\geq g(k)\}}\right] \\
\leq\ & 3\sum_{m=n}^{\infty} \frac{K(g(m))}{\alpha(g(m))} e^{-\frac{m}{2}\alpha(g(m))\lambda(g(m))} E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0)\right] \\
+\ & 6\sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha^2(g(k+1))\lambda(g(k+1))} E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0) 1_{\{|M_k|\geq g(k)\}}\right] \\
\leq\ & 3\sum_{m=n}^{\infty} \frac{K(g(m))}{\alpha(g(m))} e^{-\frac{m}{2}\alpha(g(m))\lambda(g(m))} E\left[\rho_1(\mu_0, Q(Y_{-m})\mu_0)\right] \\
+\ & 6\sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha^2(g(k+1))\lambda(g(k+1))} E^{1/2}\left[\rho_1^2(\mu_0, Q(Y_{-m})\mu_0)\right] P^{1/2}(|M_k|\geq g(k)) \\
\leq\ & 3E\left[\rho_1(\mu_0, Q(Y_0)\mu_0)\right] \sum_{m=n}^{\infty} \frac{K(g(m))}{\alpha(g(m))} e^{-\frac{m}{2}\alpha(g(m))\lambda(g(m))} \\
+\ & 6E^{1/2}\left[\rho_1^2(\mu_0, Q(Y_0)\mu_0)\right] \sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha^2(g(k+1))\lambda(g(k+1))} \sqrt{\ell(k)},
\end{aligned}
$$

where we have used the closed form expression for the sum of geometric series, the Cauchy inequality and the fact that the law of $\rho_1(\mu_0, Q(Y_{-m})\mu_0)$ equals that of $\rho_1(\mu_0, Q(Y_0)\mu_0)$.

Noting that $\rho_1(\mu_0, \mu_1) < \infty$ by Assumption 2.6, it follows from $r_1(0) + r_2(0) < \infty$ that

$$\sum_{n=0}^{\infty} \rho_1(\mu_n, \mu_{n+1}) < \infty,$$

so μ_n, $n \geq 0$ is a Cauchy sequence for the complete metric ρ_1. Hence it converges to some probability μ_* as $n \to \infty$. The claimed convergence rate also follows by the above estimates. □

Proof of Theorem 2.13. Estimates of Theorem 2.11 and (21) imply

$$\rho_0(\mu_n(\mathbf{y}_n), \mu_{n+1}(\mathbf{y}_{n+1})) \leq (1-\alpha(\bar{y}_n)\lambda(\bar{y}_n)/2)^n \rho_1(\mu_0, Q(y_{-n})\mu_0).$$

This leads to

$$
\begin{aligned}
\rho_0(\mu_n, \mu_{n+1}) \leq E[\rho_0(\mu_n(\mathbf{Y}_n), \mu_{n+1}(\mathbf{Y}_{n+1}))] &\leq \\
(1-\alpha(g(n))\lambda(g(n))/2)^n E[\rho_1(\mu_0, Q(Y_{-n})\mu_0) 1_{\{M_n<g(n)\}}] + 2P(M_n \geq g(n)) &\leq \\
(1-\alpha(g(n))\lambda(g(n))/2)^n E[\rho_1(\mu_0, Q(Y_0)\mu_0)] + 2P(M_n \geq g(n)) &\leq \\
C[e^{-n\alpha(g(n))\lambda(g(n))/2} + \ell(n)],&
\end{aligned}
$$

for some $C > 0$, using (21), Assumptions 2.7 and 2.12. The result now follows as in the proof of Theorem 2.11 above. □

5 L-mixing processes

Let $\mathcal{G}_t$, $t \in \mathbb{N}$ be an increasing sequence of sigma-algebras (i.e. a discrete-time filtration) and let $\mathcal{G}_t^+$, $t \in \mathbb{N}$ be a *decreasing* sequence of sigma-algebras such that, for each $t \in \mathbb{N}$, $\mathcal{G}_t$ is independent of $\mathcal{G}_t^+$.

Let W_t, $t \in \mathbb{N}$ be a real-valued stochastic process. For each $r \geq 1$, introduce

$$M_r(W) := \sup_{t \in \mathbb{N}} E^{1/r}[|W_t|^r].$$

For each process W such that $M_1(W) < \infty$ we also define, for each $r \geq 1$, the quantities

$$\gamma_r(W, \tau) := \sup_{t \geq \tau} E^{1/r}[|W_t - E[W_t|\mathcal{G}_{t-\tau}^+]|^r], \ \tau \geq 1, \ \Gamma_r(W) := \sum_{\tau=1}^{\infty} \gamma_r(W, \tau).$$

For some $r \geq 1$, the process W is called *L-mixing of order* r with respect to $(\mathcal{G}_t, \mathcal{G}_t^+)$, $t \in \mathbb{N}$ if it is adapted to $(\mathcal{G}_t)_{t \in \mathbb{N}}$ and $M_r(W) < \infty$, $\Gamma_r(W) < \infty$. We say that W is *L-mixing* if it is L-mixing of order r for all $r \geq 1$. This notion of mixing was introduced in [11].

Remark 5.1. It is easy to check that if W_t, $t \in \mathbb{N}$ is L-mixing of order r then also the process $\tilde{W}_t := W_t - EW_t$, $t \in \mathbb{N}$ is L-mixing of order r, moreover, $\Gamma_r(\tilde{W}) = \Gamma_r(W)$ and $M_r(\tilde{W}) \leq 2M_r(W)$.

The next lemma (Lemma 2.1 of [11]) is useful when checking the L-mixing property for a given process.

Lemma 5.2. *Let $\mathcal{G} \subset \mathcal{F}$ be a sigma-algebra, X, Y random variables with $E^{1/r}[|X|^r] < \infty$, $E^{1/r}[|Y|^r] < \infty$ with some $r \geq 1$. If Y is $\mathcal{G}$-measurable then*

$$E^{1/r}[|X - E[X|\mathcal{G}]|^r] \leq 2E^{1/r}[|X - Y|^r]$$

holds. □

L-mixing is, in many cases, easier to show than other, better-known mixing concepts and it leads to useful inequalities like Lemma 5.3 below. For further related results, see [11].

Lemma 5.3. *For an L-mixing process W of order $r \geq 2$ satisfying $E[W_t] = 0$, $t \in \mathbb{N}$,*

$$E^{1/r}\left[\left|\sum_{i=1}^{N} W_i\right|^r\right] \leq C_r N^{1/2} M_r^{1/2}(W) \Gamma_r^{1/2}(W),$$

holds for each $N \geq 1$ with a constant C_r that does not depend either on N or on W.

Proof. This follows from Theorem 1.1 of [11]. □

6 Proofs of ergodicity

Throughout this section let the assumptions of Theorem 2.14 be valid: let $\mathcal{X}$ be a Polish space with Borel field $\mathfrak{B}$; let Assumptions 2.2 and 2.6 be in force; let Assumption 2.5 hold with $R(n) := 8K(n)/\lambda(n)$, $n \in \mathbb{N}$; assume $r_1(0) + r_2(0) < \infty$ and

$$\left(\frac{K(g(N))}{\lambda(g(N))}\right)^{2\delta} \frac{\pi(N)}{N} \to 0, \ N \to \infty.$$

We now present a construction that is crucial for proving Theorem 2.14. The random mappings T_t in the lemma below serve to provide the coupling effects that are needed for establishing the L-mixing property (see Section 5 above) for an auxiliary process (Z below) which will, in turn, lead to Theorem 2.14. Such a representation with random mappings was used in [1, 3, 13, 19]. In our setting, however, there is also dependence on $y \in \mathcal{Y}$.

For $R \geq 0$, denote by $\mathfrak{C}(R)$ the set of $\mathcal{X} \to \mathcal{X}$ mappings that are constant on $C(R) = \{x \in \mathcal{X} : V(x) \leq R\}$.

Lemma 6.1. *There exists a sequence of measurable functions $T_t : \mathcal{Y} \times \mathcal{X} \times \Omega \to \mathcal{X}$, $t \geq 1$ such that*

$$P(T_t(y, x, \omega) \in A) = Q(y, x, A), \tag{23}$$

for all $t \geq 1$, $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $A \in \mathfrak{B}$ and there are events $J_t(y) \in \mathcal{F}$, for all $t \geq 1$, $y \in \mathcal{Y}$ such that

$$J_t(y) \subset \{\omega : T_t(y, \cdot, \omega) \in \mathfrak{C}(R(\|y\|))\} \text{ and } P(J_t(y)) \geq \alpha(\|y\|). \tag{24}$$

For each $t \geq 1$, let $\mathcal{L}_t$ denote the sigma-algebra generated by the random variables $T_t(y, x, \cdot)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. These sigma-algebras are independent.

Proof. Let U_n, $n \in \mathbb{N}$ be an independent sequence of uniform random variables on $[0, 1]$. Let ε_n, $n \in \mathbb{N}$ be another such sequence, independent of $(U_n)_{n\in\mathbb{N}}$. By enlarging the probability space, if necessary, we can always construct such random variables and we may even assume that (U_n, ε_n), $n \in \mathbb{N}$ are independent of $(X_0, (Y_t)_{t\in\mathbb{Z}})$.

We assume that $\mathcal{X}$ is uncountable, the case of countable $\mathcal{X}$ being analogous, but simpler. As $\mathcal{X}$ is Borel-isomorphic to $\mathbb{R}$, see page 159 of [9], we may and will assume that, actually, $\mathcal{X} = \mathbb{R}$ (we omit the details).

The main idea in the arguments below is to separate the "independent component" $\alpha(n)\nu_n(\cdot)$ from the rest of the kernel $Q(y, x, \cdot) - \alpha(n)\nu_n(\cdot)$ for $y \in A_n$ and $x \in C(R(n))$. This independent component will ensure the existence of the constant mappings in (24).

Recall the sets A_n, $n \in \mathbb{N}$ from Assumption 2.2. Let $B_n := A_n \setminus A_{n-1}$, $n \in \mathbb{N}$, with the convention $A_{-1} := \emptyset$. For each $n \in \mathbb{N}$, $y \in B_n$, let $j_n(y, r) := \nu_n((-\infty, r])$, $r \in \mathbb{R}$ (the cumulative distribution function of ν_n) and define its ($\mathfrak{A} \otimes \mathcal{B}(\mathbb{R})$-measurable) pseudoinverse by $j_n^-(y, z) := \inf\{r \in \mathbb{Q} : j(y, r) \geq z\}$, $z \in \mathbb{R}$. Here $\mathcal{B}(\mathbb{R})$ refers to the Borel-field of $\mathbb{R}$. Similarly, for $y \in B_n$ and $x \in C(R(n))$, let

$$q(y, x, r) := \frac{Q(y, x, (-\infty, r]) - \alpha(n) j_n(y, r)}{1 - \alpha(n)}, \ r \in \mathbb{R},$$

the cumulative distribution function of the normalization of $Q(y, x, \cdot) - \alpha(n)\nu_n(\cdot)$. For $x \notin C(R(n))$, set simply

$$q(y, x, r) := Q(y, x, (-\infty, r]), \ r \in \mathbb{R}.$$

For each $x \in \mathcal{X}$, define

$$q^-(y, x, z) := \inf\{r \in \mathbb{Q} : q(y, x, r) \geq z\}, \ z \in \mathbb{R}.$$

Define, for $n \in \mathbb{N}$, $y \in B_n$,

$$\begin{aligned} T_t(y, x, \omega) &:= q^-(y, x, \varepsilon_t), \text{ if } U_t(\omega) > \alpha(n) \text{ or } U_t(\omega) \leq \alpha(n) \text{ but } x \notin C(R(n)), \\ T_t(y, x, \omega) &:= j_n^-(y, \varepsilon_t), \text{ if } U_t(\omega) \leq \alpha(n) \text{ and } x \in C(R(n)). \end{aligned}$$

Notice that $T_t(y, \cdot, \omega) \in \mathfrak{C}(R(\|y\|))$ whenever $U_t(\omega) \leq \alpha(n)$, this implies (24) with $J_t(y) := \{\omega : U_t(\omega) \leq \alpha(\|y\|)\}$. The claimed independence of the sequence of sigma-algebras clearly holds. It is easy to check (23), too. □

Remark 6.2. Note that, in the above construction, $(U_n, \varepsilon_n)_{n\in\mathbb{N}}$ was taken to be independent of $(X_0, (Y_t)_{t\in\mathbb{Z}})$. This will be important later, in the proof of Theorem 2.14.

We drop dependence of the mappings T_t on ω in the notation from now on and will simply write $T_t(y, x)$. We continue our preparations for the proof of Theorem 2.14. Let $\mathcal{G}_t := \sigma(\varepsilon_i, U_i, \ i \leq t)$ and $\mathcal{G}_t^+ := \sigma(\varepsilon_i, U_i, \ i \geq t + 1)$, $t \in \mathbb{N}$. Take an arbitrary element $\tilde{x} \in \mathcal{X}$, this will remain fixed throughout this section.

Our approach to the ergodic theorem for X does not rely on the Markovian structure, it proceeds rather through establishing a convenient mixing property. The ensuing arguments will lead to Theorem 2.14 via the L-mixing property of certain auxiliary Markov chains. It turns out that L-mixing is particularly well-adapted to Markov chains, even when they are inhomogeneous (and for us this is the crucial point). The main ideas of the arguments below go back to [1], [3], [13] and [19]. In [13] and [19], Doeblin chains were treated. We need to extend those arguments substantially in the present, more complicated setting.

Let us fix $\mathbf{y} = (y_0, y_1, \ldots) \in \mathcal{Y}^{\mathbb{N}}$ till further notice such that, for some $H \in \mathbb{N}$, $\|y_j\| \leq H$ holds for all $j \in \mathbb{N}$.

Define $Z_0 := X_0$, $Z_{t+1} := T_{t+1}(y_t, Z_t)$, $t \in \mathbb{N}$. Clearly, the process Z heavily depends on the choice of $\mathbf{y}$. However, for a while we do not signal this dependence for notational simplicity. Fix also $m \in \mathbb{N}$ till further notice. Define $\tilde{Z}_m := \tilde{x}$, $\tilde{Z}_{t+1} := T_{t+1}(y_t, \tilde{Z}_t)$, $t \geq m$. Notice that $\tilde{Z}_t$, $t \geq m$ are $\mathcal{G}_m^+$-measurable.

Our purpose will be to prove that, with a large probability, $Z_{m+\tau} = \tilde{Z}_{m+\tau}$ for τ large enough. In other words, a coupling between the processes Z and $\tilde{Z}$ is realized.

Fix $\epsilon > 0$ which will be specified later. Let $\tau \geq 1$ be an arbitrary integer. Denote $\vartheta := \lceil \epsilon\tau \rceil$.

Recall that $R(H) = 8K(H)/\lambda(H)$. Define $D := C(R(H)/2) = \{x \in \mathcal{X} : V(x) \leq R(H)/2\}$ and $\overline{D} := \{(x_1, x_2) \in \mathcal{X}^2 : V(x_1) + V(x_2) \leq R(H)\}$.

Now let us notice that if $z \in \mathcal{X} \setminus D$, then for all $y \in A_H$,

$$\begin{aligned} [Q(y)(K(H)+V)](z) &\leq (1-\lambda(H))V(z) + 2K(H) \\ &\leq (1-\lambda(H)/2)V(z). \end{aligned} \tag{25}$$

Denote $\overline{Z}_t := (Z_t, \tilde{Z}_t)$, $t \geq m$. Define the $(\mathcal{G}_t)_{t\in\mathbb{N}}$-stopping times

$$\sigma_0 := m, \ \sigma_{n+1} := \min\{i > \sigma_n : \ \overline{Z}_i \in \overline{D}\}.$$

Lemma 6.3. *We have* $\sup_{k\in\mathbb{N}} E[V(Z_k)] \leq E[V(X_0)] + K(H)/\lambda(H) < \infty$. *Furthermore,* $\sup_{k\geq m} E[V(\tilde{Z}_k)] \leq V(\tilde{x}) + K(H)/\lambda(H)$.

Proof. Assumption 2.2 easily implies that, for $k \geq 1$,

$$E[V(Z_k)] \leq (1-\lambda(H))E[V(Z_{k-1})] + K(H).$$

Assumption 2.6 implies that $E[V(X_0)] = E[V(Z_0)] < \infty$ so, for every $k \in \mathbb{N}$,

$$E[V(Z_k)] \leq E[V(X_0)] + \sum_{l=0}^{\infty} K(H)(1-\lambda(H))^l = E[V(X_0)] + \frac{K(H)}{\lambda(H)}.$$

Similarly,

$$E[V(\tilde{Z}_k)] \leq V(\tilde{x}) + \sum_{l=0}^{\infty} K(H)(1-\lambda(H))^l = V(\tilde{x}) + \frac{K(H)}{\lambda(H)}.$$

□

The counterpart of the above lemma for X (driven by Y, which is stochastic) instead of Z is the following.

Lemma 6.4.

$$\sup_{n\in\mathbb{N}} E[V(X_n)] < \infty.$$

Proof. Note that $E[V(X_0)] < \infty$ by Assumption 2.6. So, for each $n \geq 1$,

$$\begin{aligned} E[V(X_n)] \leq \int_{\mathcal{X}} (1+V(z))\mu_n(dz) &\leq \\ \int_{\mathcal{X}} (1+V(z))|\mu_n - \mu_0|(dz) + \int_{\mathcal{X}} (1+V(z))\mu_0(dz) &= \\ \rho_1(\mu_n, \mu_0) + E[V(X_0)] + 1. & \end{aligned}$$

As $\rho_1(\mu_n, \mu_0) \to \rho_1(\mu_*, \mu_0)$ by Theorem 2.11, the statement follows. □

The results below serve to control the number of returns to $\overline{D}$ and the probability of coupling between the processes Z and $\tilde{Z}$. Our estimation strategy in the proof of Theorem 2.14 will be the following. We will control $P(\tilde{Z}_{\tau+m} \neq Z_{\tau+m})$ for large τ: either there were only few returns of the process $\overline{Z}$ to $\overline{D}$ (which happens with small probability) or there were many returns but coupling did not occur (which also has small probability). First let us present a lemma controlling the number of returns to $\overline{D}$.

Lemma 6.5. *There is $\bar{C} > 0$ such that*

$$\sup_{n\geq 1} E\left[\exp(\varrho(H)(\sigma_{n+1}-\sigma_n))\middle|\mathcal{G}_{\sigma_n}\right] \leq \frac{\bar{C}}{\lambda^2(H)},$$

and

$$E[\exp(\varrho(H)(\sigma_1-\sigma_0))] \leq \frac{\bar{C}}{\lambda^2(H)}$$

where $\varrho(H) := \ln(1+\lambda(H)/2)$. In particular, $\sigma_n < \infty$ a.s. for each $n \in \mathbb{N}$. Furthermore, $\bar{C}$ does not depend on either $\mathbf{y}$, m or H.

Proof. We can estimate, for $k \geq 1$ and $n \geq 1$,

$$\begin{aligned}
P(\sigma_{n+1}-\sigma_n > k|\mathcal{G}_{\sigma_n}) = P(\overline{Z}_{\sigma_n+k} \notin \overline{D}, \ldots, \overline{Z}_{\sigma_n+1} \notin \overline{D}|\mathcal{G}_{\sigma_n}) &\leq \\
E\left[\left(\frac{V(Z_{\sigma_n+k})+V(\tilde{Z}_{\sigma_n+k})}{R(H)}\right)1_{\{\overline{Z}_{\sigma_n+k-1}\notin\overline{D}\}}\cdots 1_{\{\overline{Z}_{\sigma_n+1}\notin\overline{D}\}}|\mathcal{G}_{\sigma_n}\right] &= \\
E\left[E\left[\left(\frac{V(Z_{\sigma_n+k})+V(\tilde{Z}_{\sigma_n+k})}{R(H)}\right)1_{\{\overline{Z}_{\sigma_n+k-1}\notin\overline{D}\}}|\mathcal{G}_{\sigma_n+k-1}\right]1_{\{\overline{Z}_{\sigma_n+k-2}\notin\overline{D}\}}\right. &\cdots \\
\left.\cdots 1_{\{\overline{Z}_{\sigma_n+1}\notin\overline{D}\}}|\mathcal{G}_{\sigma_n}\right]. &
\end{aligned}$$

Notice that, on $\{\overline{Z}_{\sigma_n+k-1} \notin \overline{D}\}$, either Z_{σ_n+k-1} or $\tilde{Z}_{\sigma_n+k-1}$ falls outside D. Let us assume that Z_{σ_n+k-1} does so, i.e. the estimation below is meant to take place on the set $\{Z_{\sigma_n+k-1} \notin D\}$. The other case can be treated analogously. Assumption 2.2 and the observation (25) imply that

$$\begin{aligned}
E\left[\left(\frac{V(Z_{\sigma_n+k})+V(\tilde{Z}_{\sigma_n+k})}{R(H)}\right)1_{\{\overline{Z}_{\sigma_n+k-1}\notin\overline{D}\}}|\mathcal{G}_{\sigma_n+k-1}\right] &\leq \\
\frac{1}{R(H)}[(1-\lambda(H)/2)V(Z_{\sigma_n+k-1})-K(H)] &+ \\
\frac{1}{R(H)}[(1-\lambda(H))V(\tilde{Z}_{\sigma_n+k-1})+K(H)] &\leq \\
\frac{1-\lambda(H)/2}{R(H)}[V(Z_{\sigma_n+k-1})+V(\tilde{Z}_{\sigma_n+k-1})]. &
\end{aligned}$$

This argument can clearly be iterated and leads to

$$\begin{aligned}
P(\sigma_{n+1}-\sigma_n > k|\mathcal{G}_{\sigma_n}) &\leq \\
\frac{(1-\lambda(H)/2)^{k-1}}{R(H)}E\left[V(Z_{\sigma_n+1})+V(\tilde{Z}_{\sigma_n+1})\middle|\mathcal{G}_{\sigma_n}\right] &\leq \\
\frac{(1-\lambda(H)/2)^{k-1}}{R(H)}\left[(1-\lambda(H))\left[V(Z_{\sigma_n})+V(\tilde{Z}_{\sigma_n})\right]+2K(H)\right] &\leq \\
\leq (1-\lambda(H)/2)^k, &
\end{aligned}$$

by Assumption 2.2, since $\overline{Z}_{\sigma_n} \in \overline{D}$. In the case $n = 0$, we arrive at

$$\begin{aligned}
P(\sigma_1-\sigma_0 > k) &\leq \\
E\left[(1-\lambda(H))(V(Z_m)+V(\tilde{x}))+2K(H)\right]\frac{(1-\lambda(H)/2)^{k-1}}{R(H)} &\leq \\
\left(E[V(X_0)]+\frac{1}{8}+V(\tilde{x})+\frac{\lambda(H)}{4}\right)\left(1-\frac{\lambda(H)}{2}\right)^{k-1} &
\end{aligned}$$

instead, in a similar way, by Lemma 6.3.

Now we turn from probabilities to expectations. Using $e^{\varrho(H)} \leq 2$, we can estimate, for $n \geq 1$,

$$\begin{aligned} E\left[\exp\{\varrho(H)(\sigma_{n+1}-\sigma_n)\}\big|\mathcal{G}_{\sigma_n}\right] &\leq \\ \sum_{k=0}^{\infty} e^{\varrho(H)(k+1)}\left(1-\frac{\lambda(H)}{2}\right)^k &\leq \\ 2\sum_{k=0}^{\infty}\left(1-\frac{\lambda^2(H)}{4}\right)^k &= \frac{8}{\lambda^2(H)}. \end{aligned}$$

When $n = 0$, we obtain

$$\begin{aligned} E\left[\exp\{\varrho(H)(\sigma_1-\sigma_0)\}\right] &\leq \\ \left(E[V(X_0)]+\frac{1}{8}+V(\tilde{x})+\frac{\lambda(H)}{4}\right)\left[e^{\varrho(H)}+\sum_{k=1}^{\infty} e^{\varrho(H)(k+1)}\left(1-\frac{\lambda(H)}{2}\right)^{k-1}\right] &\leq \\ \frac{\bar{C}}{\lambda^2(H)}, \end{aligned}$$

for some $\bar{C} \geq 8$. The statement follows. □

Now we make the choice

$$\epsilon := \epsilon(H) = \rho(H)/4(\ln(\bar{C}) - 2\ln(\lambda(H))).$$

Corollary 6.6. *If*

$$\tau \geq 1/\epsilon(H), \tag{26}$$

then

$$P(\sigma_\vartheta > m+\tau) \leq \exp(-\varrho(H)\tau/2).$$

Proof. Lemma 6.5 and the tower rule for conditional expectations easily imply

$$E[\exp(\varrho(H)\sigma_\vartheta)] \leq \left(\frac{\bar{C}}{\lambda^2(H)}\right)^{\vartheta} e^{\varrho(H)m}.$$

Hence, by the Markov inequality,

$$P(\sigma_\vartheta > m+\tau) \leq \left(\frac{\bar{C}}{\lambda^2(H)}\right)^{\vartheta} \exp(-\varrho(H)\tau).$$

The statement now follows by direct calculations. Indeed, this choice of $\epsilon(H)$ and $\tau \geq 1/\epsilon(H)$ imply

$$(\ln(\bar{C}) - 2\ln(\lambda(H)))[\epsilon(H)\tau + 1] \leq \frac{\tau}{2}\ln(1+\lambda(H)/2),$$

which guarantees

$$(\ln(\bar{C}) - 2\ln(\lambda(H))\lceil\epsilon(H)\tau\rceil - \tau\ln(1+\lambda(H)/2) \leq -\frac{\tau}{2}\ln(1+\lambda(H)/2).$$

□

The next lemma controls the probability of coupling between Z and $\tilde{Z}$.

Lemma 6.7.

$$P(Z_{m+\tau} \neq \tilde{Z}_{m+\tau},\ \sigma_\vartheta \leq m+\tau) \leq (1-\alpha(H))^{\vartheta-1} \leq e^{-(\vartheta-1)\alpha(H)}.$$

Proof. For typographical reasons, we will write $\sigma(n)$ instead of σ_n in this proof. Notice that if $\omega \in \Omega$ is such that $\sigma(k)(\omega) < m+\tau$ and $T_{\sigma(k)(\omega)+1}(y_{\sigma(k)(\omega)+1}, \cdot, \omega) \in \mathfrak{C}(R(H))$ then $Z_{\sigma(k)(\omega)+1}(\omega) = \tilde{Z}_{\sigma(k)(\omega)+1}(\omega)$ hence also $Z_{m+\tau}(\omega) = \tilde{Z}_{m+\tau}(\omega)$. Recall the proof of Lemma 6.1 and estimate

$$\begin{aligned} P(Z_{m+\tau} \neq \tilde{Z}_{m+\tau},\ \sigma(\vartheta) \leq m+\tau) &\leq \\ P(U_{\sigma(1)+1} > \alpha(H), \ldots, U_{\sigma(\vartheta-1)+1} > \alpha(H)) &= \\ E[E[1_{\{U_{\sigma(\vartheta-1)+1} > \alpha(H)\}} | \mathcal{G}_{\sigma(\vartheta-1)}] 1_{\{U_{\sigma(1)+1} > \alpha(H)\}} \cdots 1_{\{U_{\sigma(\vartheta-2)+1} > \alpha(H)\}}]. \end{aligned}$$

As easily seen,

$$E[1_{\{U_{\sigma(\vartheta-1)+1} > \alpha(H)\}} | \mathcal{G}_{\sigma(\vartheta-1)}] = (1 - \alpha(H)).$$

Iterating the above argument, we arrive at the statement of this lemma using $1-x \leq e^{-x}$, $x \geq 0$. $\square$

Lemma 6.8. *Let $\phi : \mathcal{X} \to \mathbb{R}$ be measurable with*

$$|\phi(x)| \leq \tilde{C}[V^{\delta}(x) + 1],\ x \in \mathcal{X}$$

for some $0 < \delta \leq 1/2$ and $\tilde{C} > 0$. Then the process $\phi(Z_t)$, $t \in \mathbb{N}$ is L-mixing of order p with respect to $(\mathcal{G}_t, \mathcal{G}_t^+)$, $t \in \mathbb{N}$, for all $1 \leq p < 1/\delta$. Furthermore, $\Gamma_p(\phi(Z))$, $M_p(\phi(Z))$ have upper bounds that do not depend on $\mathbf{y}$, only on H.

In the sequel we will use, without further notice, the following elementary inequalities for $x, y \geq 0$:

$$(x+y)^r \leq 2^{r-1}(x^r + y^r) \text{ if } r \geq 1;\ (x+y)^r \leq x^r + y^r \text{ if } 0 < r < 1.$$

Proof of Lemma 6.8. Clearly,

$$M_{1/\delta}(\phi(Z)) \leq \tilde{C}\left[1 + \left(E[V(X_0)] + \frac{K(H)}{\lambda(H)}\right)^{\delta}\right],$$

by Lemma 6.3. Also,

$$M_p(\phi(Z)) \leq M_{1/\delta}(\phi(Z)),$$

for all $1 \leq p < 1/\delta$.

Now we turn to establishing a bound for $\Gamma_p(\phi(Z))$. Since $\tilde{Z}_m$ is deterministic, $\tilde{Z}_{m+\tau}$ is $\mathcal{G}_m^+$-measurable. Lemma 5.2 implies that, for $\tau \geq 1$,

$$\begin{aligned} E^{1/p}[|\phi(Z_{m+\tau}) - E[\phi(Z_{m+\tau})|\mathcal{G}_m^+]|^p] &\leq \\ 2E^{1/p}[|\phi(Z_{m+\tau}) - \phi(\tilde{Z}_{m+\tau})|^p] &\leq \\ 2E^{1/p}[(|\phi(Z_{m+\tau})| + |\phi(\tilde{Z}_{m+\tau})|)^p 1_{\{Z_{m+\tau} \neq \tilde{Z}_{m+\tau}\}}] &\leq \\ 2E^{\delta}[(|\phi(Z_{m+\tau})| + |\phi(\tilde{Z}_{m+\tau})|)^{1/\delta}] P^{\frac{1-p\delta}{p}}(Z_{m+\tau} \neq \tilde{Z}_{m+\tau}), \end{aligned} \tag{27}$$

using Hölder's inequality with the exponents $1/(p\delta)$ and $1/(1-p\delta)$. By Lemma 6.3,

$$\begin{aligned} E^{\delta}[(|\phi(Z_{m+\tau})| + |\phi(\tilde{Z}_{m+\tau})|)^{1/\delta}] &\leq \\ \tilde{C}\left[1 + \left(E[V(X_0)] + \frac{K(H)}{\lambda(H)}\right)^{\delta}\right] &+ \\ \tilde{C}\left[1 + \left(V(\tilde{x}) + \frac{K(H)}{\lambda(H)}\right)^{\delta}\right] &\leq \check{C}\left[\frac{K(H)}{\lambda(H)}\right]^{\delta}, \end{aligned} \tag{28}$$

for some suitable $\check{C} > 0$. Since

$$P(Z_{m+\tau} \neq \tilde{Z}_{m+\tau}) \leq P(Z_{m+\tau} \neq \tilde{Z}_{m+\tau},\ \sigma_{\vartheta} \leq m+\tau) + P(\sigma_{\vartheta} > m+\tau),$$

we obtain from Lemma 6.7 and Corollary 6.6 that for τ satisfying (26),

$$\begin{aligned}&\gamma_p(\phi(Z),\tau)\\ \leq\ & 2\check{C}\left(\frac{K(H)}{\lambda(H)}\right)^{\delta}\left[\exp\left(-\alpha(H)[\epsilon(H)\tau-1](1-p\delta)/p\right)+\exp\left(-\frac{\varrho(H)\tau}{2}(1-p\delta)/p\right)\right],\end{aligned}$$

noting that the estimates of Lemma 6.7 and Corollary 6.6 do not depend on the choice of m. For each integer

$$1\leq\tau<1/\epsilon(H),$$

we will apply the trivial estimate

$$\gamma_p(\phi(Z),\tau)\leq 2M_p(\phi(Z))\leq 2M_{1/\delta}(\phi(Z))\leq 2\check{C}\left[\frac{K(H)}{\lambda(H)}\right]^{\delta},$$

recall (28). Hence

$$\begin{aligned}\Gamma_p(\phi(Z))\leq 2\check{C}\frac{1}{\epsilon(H)}\left(\frac{K(H)}{\lambda(H)}\right)^{\delta} &+\\ \sum_{\tau\geq 1/\epsilon(H)}\left[\exp\left(-\alpha(H)[\epsilon(H)\tau-1](1-p\delta)/p\right)+\exp\left(-\frac{\varrho(H)\tau}{2}(1-p\delta)/p\right)\right]\left(\frac{K(H)}{\lambda(H)}\right)^{\delta} &\leq\\ c'\left[\frac{1}{\epsilon(H)}+\frac{\exp\left(\alpha(H)(1-p\delta)/p\right)}{1-\exp\left(-\alpha(H)\epsilon(H)(1-p\delta)/p\right)}+\frac{1}{1-\exp\left(-\frac{\varrho(H)(1-p\delta)}{2p}\right)}\right]\left(\frac{K(H)}{\lambda(H)}\right)^{\delta} &\leq\\ c''\left[\frac{1}{\alpha(H)\epsilon(H)}+\frac{1}{\lambda(H)}\right]\left(\frac{K(H)}{\lambda(H)}\right)^{\delta} &\leq\\ c'''\frac{|\ln(\lambda(H))|}{\alpha(H)\lambda(H)}\left(\frac{K(H)}{\lambda(H)}\right)^{\delta} &\qquad (29)\end{aligned}$$

with some $c',c'',c'''>0$, using elementary properties of the functions $x\to 1/(1-e^{-x})$ and $x\to\ln(1+x)$. The L-mixing property of order p follows. (Note, however, that c''' depends on p, δ as well as on $E[V(X_0)]$.) □

Proof of Theorem 2.14. Now we start signalling the dependence of Z on $\mathbf{y}$ and hence write $Z_t^{\mathbf{y}}$, $t\in\mathbb{N}$. For each $\mathbf{y}\in\mathcal{Y}^{\mathbb{N}}$, define $W_t(\mathbf{y}):=\phi(Z_t^{\mathbf{y}})-E[\phi(Z_t^{\mathbf{y}})]$, $t\in\mathbb{N}$. Let $\mathbf{Y}\in\mathcal{Y}^{\mathbb{N}}$ be defined by $\mathbf{Y}_j=Y_j$, $j\in\mathbb{N}$. Note that the law of $Z_t^{\mathbf{Y}}$, $t\in\mathbb{N}$ equals that of X_t, $t\in\mathbb{N}$, by construction of Z and by Remark 6.2.

Fix $p\geq 2$. Fix $N\in\mathbb{N}$ for the moment. In the particular case where $\mathbf{y}$ satisfies $|y_j|\leq g(N)$, $j\in\mathbb{N}$, the process $W_t(\mathbf{y})$, $t\in\mathbb{N}$ is L-mixing by Lemma 6.8 and Remark 5.1. Hence Lemma 5.3 implies

$$\begin{aligned}E^{1/p}\left[\left|\frac{W_1(\mathbf{y}_1)+\ldots+W_N(\mathbf{y}_N)}{N}\right|^p\right] &\leq\\ \frac{C_pM_p^{1/2}(W(\mathbf{y}))\Gamma_p^{1/2}(W(\mathbf{y}))}{N^{1/2}} &\leq\\ \frac{C_pM_{1/\delta}^{1/2}(W(\mathbf{y}))\Gamma_p^{1/2}(W(\mathbf{y}))}{N^{1/2}} &\leq\\ \frac{2C_p\sqrt{\check{C}}[K(g(N))/\lambda(g(N))]^{\delta/2}\sqrt{c'''}[K(g(N))/\lambda(g(N))]^{\delta/2}\pi^{1/2}(N)}{N^{1/2}},\end{aligned}$$

by (28) and (29); recall also Remark 5.1. Fix $\tilde{y}\in A_0$ and define

$$\tilde{Y}_j:=Y_j,\ \text{if}\ Y_j\in A_{g(N)},\ \tilde{Y}_j:=\tilde{y},\ \text{if}\ Y_j\notin A_{g(N)}.$$

Note that, by (10),

$$E^{\delta}[|W_j(Y_j)|^{1/\delta}]\leq 2\tilde{C}(1+E^{\delta}[V(X_j)]),\ j\geq 1.$$

Estimate, using Hölder's inequality with exponents $1/(\delta p)$, $1/(1-\delta p)$,

$$
\begin{aligned}
E^{1/p}\left[\left|\frac{(\phi(X_1)-E[\phi(X_1)])+\ldots+(\phi(X_N)-E[\phi(X_N)])}{N}\right|^p\right] &= \\
E^{1/p}\left[\left|\frac{W_1(Y_1)+\ldots+W_N(Y_N)}{N}\right|^p\right] &\leq \\
E^{1/p}\left[\left|\frac{W_1(\tilde{Y}_1)+\ldots+W_N(\tilde{Y}_N)}{N}\right|^p\right] &+ \\
M_{1/\delta}(W(Y))P^{\frac{1-p\delta}{p}}((\tilde{Y}_1,\ldots,\tilde{Y}_N)\neq(Y_1,\ldots,Y_N)) &\leq \\
\frac{C'[K(g(N))/\lambda(g(N))]^{\delta}\pi^{1/2}(N)}{N^{1/2}}+C'\left(1+\sup_{n\in\mathbb{N}}E[V(X_n)]\right)^{\delta}\ell^{\frac{1-p\delta}{p}}(N) &\leq \\
\frac{C''[K(g(N))/\lambda(g(N))]^{\delta}\pi^{1/2}(N)}{N^{1/2}}+C''\ell^{\frac{1-p\delta}{p}}(N), & \qquad (30)
\end{aligned}
$$

with some constants $C', C''>0$, using Lemma 6.4.

Since $E[\phi(Z_t^{\mathbf{Y}})]=E[\phi(X_t)]$ converges to $\int_{\mathcal{X}}\phi(x)\,\mu_*(dx)$ at the rate given by Theorem 2.11, we can conclude that L^p convergence of the averages indeed takes place. More precisely,

$$
\begin{aligned}
E^{1/p}\left|\frac{\phi(X_1)+\ldots+\phi(X_N)}{N}-\int_{\mathcal{X}}\phi(z)\mu_*(dz)\right|^p &\leq \\
C\left[\sqrt{\frac{\pi(N)[K(g(N))/\lambda(g(N))]^{2\delta}}{N}}+\ell^{\frac{1-p\delta}{p}}(N)+\frac{\sum_{j=1}^N[r_1(j)+r_2(j)]}{N}\right] & \qquad (31)
\end{aligned}
$$

holds for $N\geq 1$, with some $C=C(p)>0$. The case $p=2$ implies directly the result for $1\leq p<2$, too. □

Proof of Theorem 2.15. This follows very closely the proof of Theorem 2.14, we only point out the differences. Denote by S an upper bound for $|\phi|$. Take an arbitrary $p\geq 2$. We may use the Hölder inequality with exponents 1 and ∞ in the estimates (27). This leads to

$$
\Gamma_p(\phi(Z))\leq c'''\frac{|\ln(\lambda(H))|}{\alpha(H)\lambda(H)},
$$

using the argument of (29). Then the proof of L^p convergence can be completed as above. Note that, instead of

$$
M_{1/\delta}(W(Y))P^{\frac{1-p\delta}{p}}((\tilde{Y}_1,\ldots,\tilde{Y}_N)\neq(Y_1,\ldots,Y_N))
$$

we may write

$$
SP((\tilde{Y}_1,\ldots,\tilde{Y}_N)\neq(Y_1,\ldots,Y_N))\leq S\ell(N)
$$

in (30). Finally, we arrive at

$$
\begin{aligned}
E^{1/p}\left|\frac{\phi(X_1)+\ldots+\phi(X_N)}{N}-\int_{\mathcal{X}}\phi(z)\mu_*(dz)\right|^p &\leq \\
C\left[\sqrt{\frac{\pi(N)}{N}}+\frac{\sum_{j=1}^N[r_3(j)+r_4(j)]}{N}\right], & \qquad (32)
\end{aligned}
$$

for some $C=C(p)>0$, noting that, since $\ell(N)\leq[\sum_{j=1}^N r_4(j)]/N$, the term containing $\ell(N)$ is subsumed in the convergence rate in (32). □

7 Appendix

If Assumptions 2.2 and 2.5 hold with *constants* λ, α, K then convergence to the limiting law takes place at a geometric rate. More precisely, the following is true.

Theorem 7.1. *Let Assumption 2.2 be in force with constants* λ, K*, and let* $R := 4K/\lambda$ *and* $C(R) := \{x \in \mathcal{X} : V(x) \leq R\}$*. Assume that there is a probability* ν *on* $\mathfrak{B}$ *such that*

$$Q(y, x, A) \geq \alpha\nu(A), \ A \in \mathfrak{B}, \ x \in C(R), \ y \in \mathcal{Y}.$$

Let Assumption 2.12 hold. Then there is a probability μ_* *on* $\mathcal{X}$ *such that*

$$\rho_1(\mu_n, \mu_*) \leq c_1 e^{-c_2 n}, \ n \in \mathbb{N},$$

with some constants $c_1, c_2 > 0$*.*

Theorem 7.2. *Let* $\mathcal{X}$ *be a Polish space and let* $\mathfrak{B}$ *be its Borel field. Let the assumptions of Theorem 7.1 hold, but with* $R := 8K/\lambda$*. For each measurable* $\phi : \mathcal{X} \to \mathbb{R}$ *satisfying* (10) *with a certain* $0 < \delta \leq 1/2$*, one has, for all* $1 \leq p < 1/\delta$*,*

$$E^{1/p} \left| \frac{\phi(X_1) + \ldots + \phi(X_N)}{N} - \int_{\mathcal{X}} \phi(z)\mu_*(dz) \right|^p \leq \frac{C}{\sqrt{N}}, \tag{33}$$

for some $C = C(p) > 0$*. If* ϕ *is bounded then* (33) *holds for each* $p \geq 1$ *and also*

$$\frac{\phi(X_1) + \ldots + \phi(X_N)}{N} \to \int_{\mathcal{X}} \phi(z)\mu_*(dz), \ N \to \infty,$$

almost surely.

Proof of Theorem 7.1. Analogously to the proof of Theorem 2.11, we obtain

$$\rho_1(\mu_n(\mathbf{y}_n), \mu_{n+1}(\mathbf{y}_{n+1})) \leq \frac{3K}{\alpha} \left(1 - \frac{\alpha\lambda}{2}\right)^n \rho_1(\mu_0, Q(-y_n)\mu_0),$$

which leads to

$$\rho_1(\mu_n, \mu_{n+1}) \leq Ce^{-n\alpha\lambda/2}, \ n \geq 1.$$

The result follows again as in the proof of Theorem 2.11 above. □

Proof of Theorem 7.2. The steps in the proofs of Theorems 2.14 and 2.15 can be repeated with λ, K, α, ϱ not depending on H. Hence π can also be chosen constant. The dominant term in (31) is of the order $1/\sqrt{N}$.

Now we turn to the proof of almost sure convergence. Take $p > 2$ and $q := (p-2)/(4p)$. Apply Markov's inequality and (30) to obtain

$$P\left(\left| \frac{W_1(Y_1) + \ldots + W_N(Y_N)}{N} \right| \geq \frac{1}{N^q} \right) \leq \frac{E\left|\frac{W_1(Y_1)+\ldots+W_N(Y_N)}{N}\right|^p}{N^{-qp}} \leq \frac{F^p(p)N^{qp}}{N^{p/2}}.$$

Almost sure convergence follows by the Borel-Cantelli lemma since

$$(p/2) - qp > 1.$$

□

Remark 7.3. Under the conditions of Theorem 7.2 we get the L-mixing property of order p for $\phi(Z^{\mathbf{y}})$, with $\Gamma_p(\phi(Z^{\mathbf{y}}))$, $M_p(\phi(Z^{\mathbf{y}}))$ admitting an upper bound independent of $\mathbf{y}$, for $1 \leq p < 1/\delta$. When ϕ is bounded, the same holds for each $p \geq 1$.

Remark 7.4. Let X_t, $t \in \mathbb{N}$ be a $\mathcal{X}$-valued *Markov chain* with $X_0 = x_0$, where $\mathcal{X}$ is a Polish space with Borel field $\mathfrak{B}$. Denoting the transition kernel of X by $Q(x, A)$, $x \in \mathcal{X}$, $A \in \mathfrak{B}$, we impose two standard assumptions (see [18, 14]) for geometric ergodicity:

$$[QV](x) \leq (1 - \lambda)V(x) + K, \ x \in \mathcal{X},$$

for some measurable function $V : \mathcal{X} \to [0, \infty)$, $0 < \lambda \leq 1$, $K > 0$ and

$$\inf_{x \in C} Q(x, A) \geq \alpha \nu(A), \;\; A \in \mathfrak{B},$$

for some probability ν, constant $\alpha > 0$ and

$$C := \{x \in \mathcal{X} : V(x) \leq 4K/\lambda\}.$$

Under these assumptions, the process X fits our framework above (choosing $\mathcal{Y}$ to be a singleton) and the arguments of Lemma 6.8 show that, for $0 < \delta \leq 1/2$ and for any measurable $\phi : \mathcal{X} \to \mathbb{R}$ satisfying

$$|\phi(x)| \leq c(1 + V^{\delta}(x)), \;\; x \in \mathcal{X},$$

with some $c > 0$, the process $\phi(X_t)$ is L-mixing of order p for each $1 \leq p < 1/\delta$. Furthermore,

$$M_p(\phi(X)) + \Gamma_p(\phi(X)) \leq \bar{c}[1 + V^{\delta}(x_0)]$$

for some $\bar{c} = \bar{c}(p) > 0$. When ϕ is bounded, the same holds for each $p \geq 1$.

Although this result forms a very particular case of our framework, it is still of interest: on one hand, it establishes a useful mixing property for a wide class of Markov processes; on the other hand, it underlines the versatility of the concept of L-mixing.

References

[1] R. Bhattacharya and M. Majumdar. On a theorem of Dubins and Freedman. *J. Theor. Probab.*, 12:1067–1087, 1999.

[2] R. N. Bhattacharya and E. C. Waymire. *Stochastic Processes with Applications.* Wiley & Sons, New York, 1990.

[3] 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.

[4] A. A. Borovkov. *Egodicity and stability of stochastic processes.* Wiley & Sons, New York, 1998.

[5] R. Cogburn. The ergodic theory of Markov chains in random environments. *Z. Wahrsch. Verw. Gebiete*, 66:109–128, 1984.

[6] R. Cogburn. On direct convergence and periodicity for transitions probabilities of Markov chains in random environments. *Ann. Probab.*, 18:642–654, 1990.

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

[8] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[9] C. Dellacherie and P.-A. Meyer. *Probability and potential.* North-Holland, Amsterdam, 1979.

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

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

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

[13] 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.

[14] 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 F.)*, Progress in Probability, vol. 63, 109–117, 2011.

[15] M. Hairer, J. Mattingly and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149:223–259, 2011.

[16] Y. Kifer. Perron-Frobenius theorem, large deviations, and random perturbations in random environments. *Math. Zeitschrift*, 222:677–698, 1996.

[17] Y. Kifer. Limit theorems for random transformations and processes in random environments. *Trans. American Math. Soc.*, 350:1481–1518, 1998.

[18] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Springer-Verlag, 1993.

[19] G. Molnár-Sáska. *Statistical analysis of hidden Markov models.* PhD thesis, Eötvös Loránd University, 2005.

[20] S. Orey. Markov chains with stochastically stationary transition probabilities. *Ann. Probab.*, 19:907–928, 1991.

[21] T. Seppäläinen. Large deviations for Markov chains with random transitions. *Ann. Probab.*, 22:713–748, 1994.

[22] Ö. Stenflo. Markov chains in random environments and random iterated function systems. *Trans. American Math. Soc.*, vol. 353, 3547–3562, 2001.

[23] B. S. Tsirelson. A geometric approach to maximum likelihood estimation for infinite-dimensional Gaussian location. II. *Theory Probab. Appl.*, 30:820–828, 1985.

[24] R. A. Vitale. The Wills functional and Gaussian processes. *Ann. Probab.*, 24:2172–2178, 1996.