On the Ergodicity of Certain Markov Chains in Random Environments

Balázs Gerencsér · Miklós Rásonyi

Received: 26 October 2020 / Revised: 13 March 2023 / Accepted: 13 March 2023 / Published online: 6 May 2023
© The Author(s) 2023

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 “drift” and “small set” 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 intractable. In particular, our results are pertinent to difference equations modulated by a stationary (Gaussian) process. Such equations arise in applications such as discretized stochastic volatility models of mathematical finance.

Keywords Ergodicity · Markov chain · Random environment

Mathematics Subject Classification (2020) 60J05 · 60K37 · 37A25

1 Introduction
Markov chains in random environments (recursive chains in the terminology of [3]) were systematically studied on countable state spaces in [4, 5, 17]. However, papers on the ergodic properties of such processes on a general state space are scarce and require rather strong, Doeblin-type conditions, see [13, 14, 18]. An exception is [19], where the system dynamics is assumed to be contracting instead. This is also rather restrictive an assumption, and only weak convergence of the laws can be 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 [16], and Foster–Lyapunov-type “drift” conditions, see Chapter 15 of [16].

Assuming that a suitably defined maximal process of the random environment satisfies a tail estimate, we manage to establish stochastic stability. We use certain ideas of [12] to obtain convergence to a limiting distribution in total variation norm with estimates on the convergence rate, see Sect. 2 for the statements of our results. We also present a method to prove ergodic theorems, exploiting ideas of [2, 11]. An important technical ingredient is the notion of \( L \)-mixing, see Sect. 5.

As examples, we present difference equations modulated by Gaussian processes in Sect. 3. These can be regarded as discretizations of diffusions in random environments which arise, for instance, in stochastic volatility models of mathematical finance, see [6, 9]. These examples demonstrate the power of our approach. Proofs appear in Sects. 4, 6 and 8. Certain ramifications are explored in Sect. 7.

\section{2 Main Results}

Let \( \mathcal{Y} \) be a Polish space with its Borel sigma-field \( \mathcal{A} \), and let \( Y_t, t \in \mathbb{Z} \) be a (strongly) stationary \( \mathcal{Y} \)-valued process on some probability space \((\Omega, \mathcal{F}, P)\). Expectation of a real-valued random variable \( X \) with respect to \( P \) will be denoted by \( E[.] \) 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 Polish space \( \mathcal{X} \) with its Borel sigma-field \( \mathcal{B} \) and denote by \( \mathcal{P}(\mathcal{X}) \) the set of probability measures on \( \mathcal{B} \). Let \( Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \to [0, 1] \) be a family of transition kernels parametrized by \( y \in \mathcal{Y} \), i.e. for all \( A \in \mathcal{B} \), \( Q(\cdot, \cdot, A) \) is \( \mathcal{A} \otimes \mathcal{B} \)-measurable and for all \( y \in \mathcal{Y}, x \in \mathcal{X}, A \to Q(y, x, A) \) is a probability on \( \mathcal{B} \). Let \( X_t, t \in \mathbb{N} \) be a \( \mathcal{X} \)-valued stochastic process such that

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

where the filtration is defined by

\[
\mathcal{F}_t := \sigma(Y_j, \ j \in \mathbb{Z}; \ X_j, \ 0 \leq j \leq t), \ t \geq 0.
\]

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 \( t \). Thus, \( X \) is a \textit{Markov chain in a random environment}. Our purpose is to study the ergodic properties of \( X \).

\textbf{Remark 2.1} Obviously, the law of \( X_t, t \in \mathbb{N} \) (and also its joint law with \( Y_t, t \in \mathbb{Z} \)) is uniquely determined by (1). For every given \( Q \) and \( X_0 \), there exists a process \( X \) satisfying (1) (after possibly enlarging the probability space). See, for example, page 228 of [1] for a similar construction. We will establish a more precise result in Lemma 6.1, under additional assumptions.

We will now introduce a number of assumptions of various kinds that will figure in the statements of the main results: Theorems 2.12, 2.16, 2.17 and 2.18.
The following assumption closely resembles the well-known drift conditions for geometrically ergodic Markov chains, see, for example, Chapter 15 of [16]. In our case, however, they are relaxed by also having dependence on the state of the random environment.

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

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

We assume that there are 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 $K(\cdot) \geq 1$.

We 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 (i.e. $\lambda(\|Y\|)$ decreases and $K(\|y\|)$ increases).

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

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

where $\lceil\cdot\rceil$ stands for the ceiling function. In the examples of the present paper, we will always have $\mathcal{B} = \mathbb{R}^d$ with some $d \geq 1$ and $|\cdot| = \|\cdot\|_\mathcal{B}$ will denote the respective Euclidean norm. Note, however, that in general $\|\cdot\|$ is not necessarily related to any geometric structure.

**Remark 2.4** It would be desirable to relax Assumption 2.2 by allowing $\lambda$ to vary in $(-\infty, 1)$ as long as “in the average” it is contractive. (There are multiple options for the precise formulation of this property.) Such a result has been worked out in [15].

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 $\nu_n$ on $\mathcal{B}$ such that, for all $y \in A_n, x \in C(R(n))$ and $A \in \mathcal{B}$,

$$Q(y, x, A) \geq \alpha(n)\nu_n(A). \tag{3}$$
In other words, if the state $y$ of the random environment is in $A_n$, we work on the set $C(4K(n)/\lambda(n))$ on which we are able to benefit from a “coupling effect” of strength $\alpha(n)$.

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

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

by setting

$$
\rho_\beta(v_1, v_2) := \int_{\mathcal{X}} [1 + \beta V(x)]|v_1 - v_2|(dx), \ v_1, v_2 \in \mathcal{P}_V(\mathcal{X}),
$$

for each $0 \leq \beta \leq 1$. Here, $|v_1 - v_2|$ is the total variation of the signed measure $v_1 - v_2$. Note that $\rho_0$ is just the total variation distance (and it can be defined for all $v_1, v_2 \in \mathcal{P}(\mathcal{X})$), while $\rho_1$ is the $(1 + V)$-weighted total variation distance.

**Definition 2.6** For a measurable $f : \mathcal{X} \to \mathbb{R}_+$, we define $\Phi_1(f)$ to be the set of measurable $\phi : \mathcal{X} \to \mathbb{R}$ such that $|\phi(z)| \leq C(1 + f(z))$, $z \in \mathcal{X}$ holds for some constant $C = C(\phi)$. Hence, $\Phi(1)$ denotes the set of bounded, measurable functions on $\mathcal{X}$.

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

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

Consistently with these definitions, $Q(Y_n)\mu$ will refer to the action of the kernel $Q(Y_n, \cdot, \cdot)$ on $\mu$. 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 nonnegative measurable $\phi$, too.

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

**Assumption 2.7** (Moment condition on the initial values)

$$
E[V^2(X_0) + V^2(X_1)] < \infty.
$$

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 2.8 (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

\[
P \left( \max_{1 \leq i \leq t} \| Y_i \| \geq g(t) \right) \leq \ell(t), \ t \geq 1.
\] (6)

Remark 2.9 It is clear that for a given process \( Y \), several choices for the pair of functions \( g, \ell \) 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 2.10 In the setting of Example 2.3, let \( Y \) be a Gaussian process in \( Y = \mathbb{R}^d \). Assumption 2.8 holds, for instance, with \( g(t) \sim \sqrt{t} \) and \( \ell(t) \) exponentially decreasing, see Sect. 3 for more details.

Remark 2.11 One can derive estimates like (6) 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

\[
E^{1/q} \left[ \max_{1 \leq i \leq t} |Y_i|^q \right] \leq E^{1/2q} \left[ \max_{1 \leq i \leq t} |Y_i|^{2q} \right] \\
\leq E^{1/2q} \left[ \sum_{i=1}^{t} |Y_i|^{2q} \right] \leq C(q)t^{1/2},
\]

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

\[
P \left( \max_{1 \leq i \leq t} |Y_i| \geq t \right) \leq \frac{C^q(q)t^{1/2}}{t^q} \leq \frac{C^q(q)}{t^{q-1/2}}.
\] (7)

Actually, for arbitrarily small \( \chi > 0 \) and arbitrarily large \( r \geq 1 \), we can set \( q = \frac{r}{\chi} + \frac{1}{2} \) in (7) and then Assumption 2.8 holds with

\[
g(t) = \lceil r^X \rceil \text{ and } \ell(t) = \frac{C^q(q)}{t^r}, \ t \geq 1,
\]

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 also Example 2.20.

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

\[
\zeta(t) := \min\{\alpha(t), \lambda(t)\}, \\
r_1(t) := \sum_{k=t}^{\infty} \frac{K(g(k))}{\alpha(g(k))} e^{-k\zeta(g(k))/4},
\]
\[ r_2(t) := \sum_{k=t}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))\xi(g(k+1))} \sqrt{\ell(k)}, \]
\[ r_3(t) := \sum_{k=t}^{\infty} e^{-k\xi(g(k))/4}, \]
\[ r_4(t) := \sum_{k=t}^{\infty} \ell(k), \]
\[ \pi(t) := \frac{|\ln(\lambda(g(t)))|}{\alpha(g(t))\lambda(g(t))}. \]

Now comes the first main result of the present paper: assuming our conditions on drift, minorization, initial values and control of the maxima, \( \text{Law}(X_t) \) will tend to a limiting law as \( t \to \infty \), provided that \( r_1(0) \) and \( r_2(0) \) are finite.

**Theorem 2.12** Define \( \mu_t := \text{Law}(X_t), t \in \mathbb{N} \). Let Assumptions 2.2, 2.5, 2.7 and 2.8 be in force. Assume

\[ r_1(0) + r_2(0) < \infty. \]  

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

\[ \rho_1(\mu_t, \mu_*) \leq C[r_1(t) + r_2(t)], \quad t \in \mathbb{N}, \]

for some constant \( C > 0 \). The limit \( \mu_* \) does not depend on \( X_0 \).

**Remark 2.13** When \( \lambda, K, \alpha \) are constant and do not depend on \( Y \), we retrieve the familiar exponential convergence rate in Theorem 2.12. Indeed, in this case we may suppose that \( A_n = A_0 = Y \) for all \( n \) and hence we may choose \( g(t) = 1 \) and \( l(t) = 0 \), for all \( t \).

Theorem 2.16 is just a variant of Theorem 2.12: relaxing the assumptions it provides convergence in a weaker sense.

**Assumption 2.14** (Weaker moment condition on the initial values)

\[ E[V(X_0) + V(X_1)] < \infty. \]

**Remark 2.15** As one of our referees pointed out, a simple sufficient condition for (9) can be formulated in terms of \( X_0, Y_0 \) only. Let Assumption 2.2 be in force. Then, \( E[V(X_1)] \leq E[V(X_0)] + E[K(\|Y_0\|)] \); hence, a sufficient condition for (9) is \( E[V(X_0) + K(\|Y_0\|)] < \infty \).

Similarly, let

\[ \int_{\mathcal{X}} V^2(z) Q(y, x, dz) \leq (1 - \lambda(\|y\|))^2 V^2(x) + K^2(\|y\|) \]

\( \square \) Springer
hold instead of (2). This implies (2) and also \( E[V^2(X_1)] \leq E[V^2(X_0)] + E[K^2(||Y_0||)] \); thus, a sufficient condition for (5) in terms of \( X_0, Y_0 \) only is \( E[V^2(X_0)] + K^2(||Y_0||) < \infty \) in this case.

**Theorem 2.16** Recall that \( \mu_t = \text{Law}(X_t), t \in \mathbb{N} \). Let Assumptions 2.2, 2.5, 2.8 and 2.14 be in force. Assume

\[
    r_3(0) + r_4(0) < \infty. \tag{10}
\]

Then, there is a probability \( \mu_* \) on \( \mathcal{X} \) such that \( \mu_t \to \mu_* \) in total variation as \( t \to \infty \). More precisely,

\[
    \rho_0(\mu_t, \mu_*) \leq C[r_3(t) + r_4(t)], \quad t \in \mathbb{N}, \tag{11}
\]

for some constant \( C > 0 \). The limit \( \mu_* \) does not depend on \( X_0 \).

Clearly, Assumption 2.7 implies Assumption 2.14 and (8) implies (10). Next, ergodic theorems corresponding to Theorems 2.12 and 2.16 are stated.

**Theorem 2.17** Let Assumptions 2.2, 2.5, 2.7 and 2.8 be in force, but with \( R(n) := 8K(n)/\lambda(n), n \in \mathbb{N} \) in Assumption 2.5. Let \( Y \) be an ergodic process. Let \( \phi \in \Phi(V^\delta) \) for some \( 0 < \delta \leq 1/2 \). Assume

\[
    r_1(0) + r_2(0) < \infty
\]

and

\[
    \left( \frac{K(g(t))}{\lambda(g(t))} \right)^{2\delta} \frac{\pi(t)}{t} \to 0, \quad t \to \infty. \tag{12}
\]

Then,

\[
    \frac{\phi(X_1) + \ldots + \phi(X_t)}{t} \to \int_{\mathcal{X}} \phi(z)\mu_*(dz), \quad t \to \infty \tag{13}
\]

holds in \( L^p \) for each \( p < 1/\delta \). (Here, \( \mu_* \) is the same as in Theorem 2.12.)

We can weaken our assumptions for bounded \( \phi \).

**Theorem 2.18** Let Assumptions 2.2, 2.5, 2.8 and 2.14 be in force, but with \( R(n) := 8K(n)/\lambda(n), n \in \mathbb{N} \) in Assumption 2.5. Let \( Y \) be an ergodic process. Assume

\[
    r_3(0) + r_4(0) < \infty
\]

and

\[
    \frac{\pi(t)}{t} \to 0, \quad t \to \infty. \tag{14}
\]

Then, for each \( \phi \in \Phi(1) \) the convergence (13) holds in \( L^p \) for all \( p \geq 1 \).
Remark 2.19 In Theorems 2.17 and 2.18, we require a slight strengthening of Assumption 2.5 by imposing (3) with $R(n) = 8K(n)/\lambda(n)$ instead of $R(n) = 4K(n)/\lambda(n)$.

Condition (14) is closely related to the condition $r_3(0) < \infty$ but none of the two implies the other. Indeed, fix $g(t) := t$. Choose $\lambda(t) := 1/2$ and $\alpha(t) := t^{\lambda(t)/\ln(t)}$, $t \geq 4$. Then, $\pi(t)/t \to 0$ but $r_3(0) = \infty$. Conversely, let $\alpha(t) := 1/2$ and $\lambda(t) := \frac{8\ln(t)}{t}$. Then, $r_3(0) < \infty$ but $\pi(t)/t$ tends to a positive constant.

Example 2.20 Let $Y$ be strongly stationary $\mathbb{R}^d$-valued with $E|Y_0|^p < \infty$ for all $p \geq 1$. 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 $c > 0$, similarly for $\lambda$). Let Assumption 2.7 hold. Then, Remark 2.11 shows (choosing $\chi$ small and $r$ large) that Theorems 2.12 and 2.17 apply.

Remark 2.21 One of our referees pointed out that more general versions of Assumptions 2.2 and 2.5 could be considered. Let $\tilde{\lambda}, \tilde{\alpha}, \tilde{K} : \mathcal{Y} \to (0, \infty)$ be measurable functions with $\tilde{K} \geq 1$ and $\tilde{\lambda}, \tilde{\alpha} \leq 1$. Instead of (2), one could assume

$$\int_{\mathcal{X}} V(z) Q(y, x, dz) \leq (1 - \tilde{\lambda}(y))V(x) + \tilde{K}(y), \ y \in \mathcal{Y}. \quad (15)$$

Instead of (3), one could assume that, for all $y \in \mathcal{Y}, x \in C(\tilde{R}(y))$ and $A \in \mathcal{B},$

$$Q(y, x, A) \geq \tilde{\alpha}(y)\nu(y, A), \quad (16)$$

with a transition kernel $\nu : \mathcal{Y} \times \mathcal{B} \to [0, 1]$. Here, $\tilde{R}(y) := 4\tilde{K}(y)/\tilde{\lambda}(y)$.

When it comes to (6), however, it is not totally clear how to formulate it in this more general setting. We only sketch one possibility here. Let $\lambda_t, \alpha_t > 0, t \in \mathbb{N}$ be non-increasing sequences, $K_t > 0, t \in \mathbb{N}$ a non-decreasing sequence. Define

$$\ell(t) := P \left( \Omega \setminus \left\{ \tilde{\lambda}(Y_j) \geq \lambda_{t-1}, \tilde{\alpha}(Y_j) \geq \alpha_{t-1}, \tilde{K}(Y_j) \leq K_{t-1}, j = 1, \ldots, t \right\} \right), \ t \geq 1. \quad (17)$$

Note that (6) is equivalent to

$$P \left( \Omega \setminus \{ Y_1, \ldots, Y_t \in A_{g(t)-1} \} \right) \leq \ell(t),$$

hence (17) looks reasonable. Redefine

$$\xi(t) := \min\{\alpha_t, \lambda_t\},$$

$$r_1(t) := \sum_{k=t}^{\infty} \frac{K_k}{\alpha_k} e^{-\kappa(k)/4},$$

$$r_2(t) := \sum_{k=t}^{\infty} \frac{K_{k+1}}{\alpha_{k+1} \xi(k + 1)} \sqrt{\ell(k)}.$$
Let (15), (16) and (5) be in force. Let us assume \( r_1(0) + r_2(0) < \infty \). Then, the conclusion of Theorem 2.12 remains true, with essentially the same proof.

### 3 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 Sect. 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 \( Y = \mathbb{R}^d \) for some \( d \) and \( X = \mathbb{R} \). We set \( \|y\| = |y| \), \( y \in Y \) as in Example 2.3. We will exclusively use \( V(x) = |x|, x \in \mathbb{R} \) in the present section.

**Remark 3.1** Let \( \xi_t, t \in \mathbb{Z} \) be a zero-mean \( \mathbb{R} \)-valued stationary Gaussian process with unit variance. Clearly,

\[
P(\max_{1 \leq i \leq t} |\xi_i| \geq t^b) \leq e^{-\frac{1}{4}t^{2b}} E \max_{1 \leq i \leq t} e^{\frac{1}{4}|\xi_i|^2} \leq e^{-\frac{1}{4}t^{2b}} \sum_{i=1}^{t} E e^{\frac{1}{4}|\xi_i|^2} \leq C e^{-\tilde{c}_1 t^{2b}} \leq C e^{-\tilde{c}_2 t^{2b}}
\]

for suitable \( C, \tilde{c}_1, \tilde{c}_2 \).

Applying these observations with \( b = 1 \) to every coordinate of \( Y \), it follows that Assumption 2.8 holds for the process \( Y \) with the choice \( g(k) = [c_1 \sqrt{k}], \ell(k) = \exp(-c_2 k) \) for some \( c_1, c_2 > 0 \) and thus \( r_4(n) \) decreases at a geometric rate as \( n \to \infty \).

More generally, choosing arbitrary \( b > 0 \), Assumption 2.8 holds for \( Y \) with the choice \( g(k) = [c_1 k^b], \ell(k) = \exp(-c_2 k^{2b}) \).

We assume throughout this section that \( \varepsilon_t, t \in \mathbb{N} \) 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 \( \varepsilon_0 \) 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, and we just try to stay as simple as possible.

**Example 3.2** First we investigate the effect of the “contraction coefficient” \( \lambda \) in (2). Let \( d := 1 \). Let \( 0 < \sigma \leq \bar{\sigma} \) be constants and \( \sigma : \mathbb{R} \times \mathbb{R} \to [\sigma, \bar{\sigma}] \) 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,

\[
f(x) \geq e^{-sx^2} / s, \quad x \geq 0,
\]

for some \( s > 0 \). We assume that the dynamics of \( X \) is given by

\[
X_0 := 0, \quad X_{t+1} := (1 - \Delta(Y_t))X_t + \sigma(Y_t, X_t)\varepsilon_{t+1}, \quad t \in \mathbb{N}.
\]
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 (11). (Note that we already have estimates for the rate $r_4(n)$ from Remark 3.1.)

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

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

Let $\eta > 0$ be arbitrary for the moment. We can estimate

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

Define the probability measures

$$\tilde{\nu}_\eta(A) := \frac{1}{2\eta} \text{Leb}(A \cap [-\eta, \eta]), \ A \in \mathcal{B}.$$ 

It follows that

$$Q(y, x, A) \geq 2\eta m(\eta) \tilde{\nu}_\eta(A), \ A \in \mathcal{B},$$

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 us now specify $\eta$ by setting $\eta := \tilde{R}(y) := 4K/\Delta(y), y \in \mathcal{Y}$ 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 $n \in \mathbb{N}$, and 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 (18) that

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

for all $A \in \mathcal{B}, y \in \mathcal{Y}, |x| \leq R([|y|])$ with some $c_3 > 0$, so (3) in Assumption 2.5 holds with

$$\alpha(n) := \frac{e^{-c_3 \tilde{R}^2(n)}}{c_3 \Delta(0)}, \ n \geq 1,$$

and $\nu_n = \tilde{\nu}_{\tilde{R}(n)}$. Now, let the function $\Delta$ 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
\[ \xi(g(k)) \geq e^{-c_4 \ln^2(k)/c_4}, \]

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

If instead of (18) we assume

\[ f(x) \geq e^{-sx}/s, \ x \geq 0, \]

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** We recall that \( f \) is assumed even, positive and non-increasing on \([0, \infty)\).

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 \( 0 < \Delta \leq 1 \) is a constant. We furthermore assume that

\[ c_5 G(y) \leq \sigma(y, x) \leq c_6 G(y), \ x \in \mathbb{R}, \]

with some even function \( G : \mathbb{R} \to (0, \infty) \) that is non-decreasing 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 \( \tilde{K}(y) = \max\{1, c_6 G(y)E[\varepsilon_0]\}, \ y \in \mathbb{R} \). Taking \( \tilde{R}(y) = 4\tilde{K}(y)/\Delta, \ y \in \mathbb{R} \), estimates as in Example 3.2 lead to

\[ \ell(y, x, A) \geq 2\tilde{R}(y)\left(\frac{2\tilde{R}(y)}{c_5 G(y)}\right)\frac{1}{c_6 G(y)}\tilde{v}_{\tilde{R}(y)}(A) \geq c_7 \tilde{v}_{\tilde{R}(y)}(A), \]

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

Assume, for example, \( G(y) \leq C[1 + |y|^q], \ y \geq 0 \) with some \( C, q > 0 \), this guarantees \( E[V^2(X_1)] = E[X_1^2] < \infty \), i.e. Assumption 2.7 holds. Choose \( g(k) = [c_1 \sqrt{k}], \ell(k) = \exp(-c_2 k), \) as discussed in Remark 3.1. Then, Theorems 2.12 and 2.17 apply.

**Example 3.4** We now investigate a discrete-time model for financial time series, inspired by the “fractional stochastic volatility model” of [6, 9].

Let \( w_t, t \in \mathbb{Z} \) and \( \varepsilon_t, t \in \mathbb{N} \) be two sequences of i.i.d. random variables such that the two sequences are also independent. Assume that \( w_t \) is 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 \( \xi_t \) describes the log-volatility of an asset in a financial market. It is reasonable to assume that \( \xi \) is a Gaussian linear process (see [9] 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 \( X_0 := 0 \),

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

with some \(-1 < \rho < 1, 0 < \Delta \leq 1\). The log-price is thus jointly driven by the noise sequences \( \varepsilon_t, w_t \). The parameter \( \Delta \) is responsible for the autocorrelation of \( X \). (\( \Delta \) is typically close to 1.) The parameter \( \rho \) controls the correlation of the price and its volatility. This is found to be nonzero (actually, negative) in empirical studies, see [7], and 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

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

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 \( K(n) = c_8 e^\xi(1 + n) \).

We now turn our attention to Assumption 2.5. Denote the density of the law of \( X_1 \) conditional on \( X_0 = x, Y_0 = (w, \xi) \) with respect to the Lebesgue measure by \( h_{x,w,\xi}(z), z \in \mathbb{R} \). Let us fix \( \eta > 0 \) for the moment. For \( x, z \in [-\eta, \eta] \), we clearly have

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

We assume from now on that \( f \), the density of \( \varepsilon_0 \), satisfies

\[
f(x) = s/(1 + x)^\chi, \quad x \geq 0
\]

with some \( s > 0, \chi > 3 \), this is reasonable as \( X_t \) has fat tails according to empirical studies, see [7]. At the same time, \( E[\varepsilon_0^2] < \infty \) and Assumption 2.7 are also 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 \( y = (w, \xi) \in \mathbb{R}^2 \). Specify \( \eta := \tilde{R}(y) \) and use (20) to obtain, as in Example 3.2,

\[
Q(y, x, A) \geq \frac{c_{10}}{(1 + |w|)^{\chi-1}} \tilde{R}(y) \mathbb{E}_{\tilde{R}(y)}(A),
\]

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

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

and $v_n = \tilde{v}_{R(n)}$. Recalling the end of Remark 3.1, and choosing $b > 0$ small enough, we can conclude that Theorems 2.12 and 2.17 apply to this stochastic volatility model.

More generally, instead of (19), we may consider

$$X_{t+1} - X_t = k(X_t) + \rho e^{\xi_t} w_t + \sqrt{1 - \rho^2} e^{\xi_t} \varepsilon_{t+1}$$

with some dissipative measurable function $k : \mathbb{R} \to \mathbb{R}$, i.e. we assume $xk(x) \leq -Ax^2 + B$ for all $x \in \mathbb{R}$ with some $A, B > 0$. Following the same steps, the applicability of Theorems 2.12 and 2.17 can be verified.

We stress that only a small fraction of relevant examples has been presented above, favouring simplicity. The results of Sect. 2 clearly apply in much greater generality.

### 4 Proofs of Stochastic Stability

Consider the $\mathcal{Y} := \mathcal{Y}^\mathbb{Z}$-valued random variable $Y := (Y_t)_{t \in \mathbb{Z}}$. By the measure decomposition theorem (see III.72 of [8]), there is a transition kernel $\tilde{\mu}_0 : \mathcal{Y} \times \mathcal{B} \to [0, 1]$ such that

$$\mu_0(A) = P(X_0 \in A) = \int_{\mathcal{Y}} \tilde{\mu}_0(y, A) \text{Law}(Y)(dy), \quad A \in \mathcal{B}. \quad (21)$$

For each $y \in \mathcal{Y}$, we will denote by $\tilde{\mu}_0(y)$ the probability $A \to \tilde{\mu}_0(y, A), A \in \mathcal{B}$ in the sequel.

Clearly, Assumption 2.7 is equivalent to

$$E \left[ \int_{\mathcal{X}} V^2(z)[\tilde{\mu}_0(Y)] + [Q(Y_0)\tilde{\mu}_0(Y)](dz) \right] < \infty. \quad (22)$$

and Assumption 2.14 is equivalent to

$$E \left[ \int_{\mathcal{X}} V(z)[\tilde{\mu}_0(Y)] + [Q(Y_0)\tilde{\mu}_0(Y)](dz) \right] < \infty. \quad (23)$$

We first recall a result which will be crucial in the arguments below.

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

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

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

$$\inf_{x \in C} L(x, A) \geq \alpha \nu(A), \ A \in \mathcal{B},$$

for some $\alpha > 0$. Then for each $\alpha_0 \in (0, \alpha)$ and for $\gamma_0 := \gamma + 2K/R$, $\rho_\beta(L\mu_1, L\mu_2) \leq \max \left\{ 1 - (\alpha - \alpha_0), \frac{2 + R\beta\gamma_0}{2 + R\beta} \right\} \rho_\beta(\mu_1, \mu_2), \ \mu_1, \mu_2 \in \mathcal{P}_V,$

holds for $\beta = \alpha_0/K$.

**Proof** See Theorem 3.1 in [12]. \qed

Next comes an easy corollary.

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

$$LV(x) \leq (1 - \lambda)V(x) + K, \ x \in \mathcal{X}, \quad (24)$$

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

$$\inf_{x \in C} L(x, A) \geq \alpha \nu(A), \ A \in \mathcal{B}, \quad (25)$$

for some $\alpha > 0$. Then,

$$\rho_\beta(L\mu_1, L\mu_2) \leq \left( 1 - \frac{\min(\alpha, \lambda)}{4} \right) \rho_\beta(\mu_1, \mu_2), \ \mu_1, \mu_2 \in \mathcal{P}_V, \quad (26)$$

holds for $\beta = \frac{\alpha}{2K}$.

**Proof** Choose $\gamma := 1 - \lambda$, and let $\alpha_0 := \alpha/2$. Note that $1 - (\alpha - \alpha_0) = 1 - \alpha/2$ and $R\beta = 2\alpha/\lambda$ holds for $\beta = \frac{\alpha}{2K}$. Also, $\gamma_0 = 1 - \lambda/2$. Applying Lemma 4.1, we estimate

$$\rho_\beta(L\mu_1, L\mu_2) \leq \max \left\{ 1 - (\alpha - \alpha_0), \frac{2 + R\beta\gamma_0}{2 + R\beta} \right\} \rho_\beta(\mu_1, \mu_2)$$

$$= \max \left\{ 1 - \alpha/2, 1 - \frac{\alpha\lambda}{2(\alpha + \lambda)} \right\} \rho_\beta(\mu_1, \mu_2).$$

Here,

$$\frac{\alpha\lambda}{2(\alpha + \lambda)} \geq \frac{\min(\alpha, \lambda) \max(\alpha, \lambda)}{4 \max(\alpha, \lambda)} \geq \frac{\min(\alpha, \lambda)}{4} \quad (27)$$

and we get the statement since $\alpha/2 \geq \frac{\min(\alpha, \lambda)}{4}$. \qed
We introduce some important notation now. If \((y, A) \rightarrow L(y, A), y \in \mathcal{Y}, A \in \mathcal{B}\) is a (not necessarily transition) kernel and \(Z\) is a \(\mathcal{Y}\)-valued random variable, then we define a measure \(E[L(Z)](\cdot)\) on \(\mathcal{B}\) via

\[
E[L(Z)](A) := E[\mathcal{L}(Z, A)], \ A \in \mathcal{B}.
\]

We will use the following trivial inequalities in the sequel:

\[
\rho_0(\cdot) \leq 2, \ \rho_0(\cdot) \leq \rho_\beta(\cdot) \leq \rho_1(\cdot) \leq \frac{1}{\beta} \rho_\beta(\cdot), \ 0 < \beta \leq 1.
\]

**Proof of Theorem 2.12.** For later use, we define the \(\mathcal{Y}\)-valued random variables \(\hat{Y}_n := (Y_{n+j})_{j \in \mathbb{Z}}, \) for each \(n \in \mathbb{Z}.\) Note that \(Y = \hat{Y}_0.\) Fix \(y := (y_j)_{j \in \mathbb{Z}} \in \mathcal{Y}\) for the moment. Set \(\tilde{y}_n := (y_{n+j})_{j \in \mathbb{Z}}, \) for each \(n \in \mathbb{Z}.\) Again, \(y = \tilde{y}_0.\) Define

\[
\mu_0(y) := \tilde{\mu}_0(y), \ \mu_n(y) := Q(y_0)Q(y_1) \cdots Q(y_{n-1})\tilde{\mu}_0(\tilde{y}_{n+1}), \ n \geq 1.
\]

Here, \(Q(y)\) is the operator acting on probabilities which is described in (4) but, instead of \(L(x, A),\) with the kernel \(Q(y, x, A).\) Fix \(n \geq 1\) and denote \(\tilde{y}_n := \max_{-n+1 \leq j \leq 0} \|y_j\|\). Since

\[
y_j \in A_{\tilde{y}_n}, \ -n + 1 \leq j \leq 0,
\]

Assumptions 2.2 and 2.5 imply that (24) and (25) hold for \(L = Q(y_j), j = -n + 1, \ldots, 0\) with \(K = K(\tilde{y}_n), \lambda = \lambda(\tilde{y}_n)\) and \(\alpha = \alpha(\tilde{y}_n)\). An \(n\)-fold application of Lemma 4.2 implies that, for \(\beta = \alpha(\tilde{y}_n)/2K(\tilde{y}_n),\)

\[
\rho_\beta(\mu_n(y), \mu_{n+1}(y)) \leq (1 - \zeta(\tilde{y}_n)/4)^n \rho_\beta(\tilde{\mu}_0(\tilde{y}_{n+1}), Q(\tilde{y}_n)\tilde{\mu}_0(\tilde{y}_{n})).
\]

By (29),

\[
\rho_1(\mu_n(y), \mu_{n+1}(y)) \leq \frac{2K(\tilde{y}_n)}{\alpha(\tilde{y}_n)} (1 - \zeta(\tilde{y}_n)/4)^n \rho_1(\tilde{\mu}_0(\tilde{y}_{n+1}), Q(\tilde{y}_n)\tilde{\mu}_0(\tilde{y}_{n})).
\]

(32)

We thus arrive at

\[
E[\rho_1(\mu_n(Y), \mu_{n+1}(Y))] \leq 2E \left[ \frac{K(M_n)}{\alpha(M_n)} (1 - \zeta(M_n)/4)^n \rho_1(\tilde{\mu}_0(\tilde{Y}_{n+1}), Q(Y_{n+1})\tilde{\mu}_0(Y_n)) \right],
\]

(33)

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 (33) separately on the events \(\{M_n \geq g(n)\}\) and \(\{M_n < g(n)\}\). Note first that, for each \(m \geq n,\)
\[
E \left[ \frac{K(M_m)}{\alpha(M_m)} \left( 1 - \frac{\zeta(M_m)}{4} \right)^m \rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1}))1_{\{M_m \geq g(m)\}} \right]
\]
\[
\leq \sum_{k=m}^{\infty} K(g(k+1)) \frac{\zeta(g(k+1))}{4} m^m
E \left[ \rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1}))1_{\{M_m \geq g(m)\}} \right]
\]
\[
\leq \sum_{k=m}^{\infty} K(g(k+1)) \left( 1 - \frac{\zeta(g(k+1))}{4} \right)^m
E \left[ \rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1}))1_{\{M_k \geq g(k)\}} \right],
\]
(34)
since \(M_k \geq M_m\). Hence, applying \(1 - x \leq e^{-x}, x \geq 0\) and (34),
\[
\sum_{m=n}^{\infty} E[\rho_1(\mu_m(Y), \mu_{m+1}(Y))]
\]
\[
\leq 2 \sum_{m=n}^{\infty} K(g(m)) \frac{\zeta(g(m))}{4} E \left[ \rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1}))1_{\{M_m \geq g(m)\}} \right]
\]
\[
+2 \sum_{m=n}^{\infty} \sum_{k=m}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))} \left( 1 - \frac{\zeta(g(k+1))}{4} \right)^m
E \left[ \rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1}))1_{\{M_k \geq g(k)\}} \right]
\]
\[
\leq 2 \sum_{m=n}^{\infty} K(g(m)) \frac{\zeta(g(m))}{4} E \left[ \rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1})) \right]
\]
\[
+8 \sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))\zeta(g(k+1))} e^{-n\zeta(g(k+1))}
E^{1/2} \left[ \rho_1^2(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1})) \right] P^{1/2}(M_k \geq g(k))
\]
\[
\leq 2 E \left[ \rho_1(\tilde{\mu}_0(\hat{Y}), Q(Y)\tilde{\mu}_0(Y)) \right] \sum_{m=n}^{\infty} \frac{K(g(m))}{\alpha(g(m))} e^{-\frac{m}{4} \zeta(g(m))}
\]
\[
+8 E^{1/2} \left[ \rho_1^2(\tilde{\mu}_0(\hat{Y}), Q(Y)\tilde{\mu}_0(Y)) \right] \sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))\zeta(g(k+1))} \sqrt{\xi(k)},
\]
(35)
where we have used the closed-form expression for the sum of geometric series and Cauchy–Schwarz in the second inequality; Assumption 2.8 and the fact that the law of \(\rho_1(\tilde{\mu}_0(\hat{Y}_1), Q(Y_0)\tilde{\mu}_0(Y))\) equals that of \(\rho_1(\tilde{\mu}_0(\hat{Y}_{m+1}), Q(Y_{m-1})\tilde{\mu}_0(\hat{Y}_{m-1}))\), for each \(m\), in the third inequality. Recall that
\[ E[\rho_1^2(\tilde{\mu}_0(Y_1), Q(Y_0)\tilde{\mu}_0(Y))] \leq 2E \left[ \int_{\mathcal{X}} (1 + V(z))^2 [\tilde{\mu}_0(Y_1) + Q(Y_0)\tilde{\mu}_0(Y)](dz) \right] \]

\[ = 2E \left[ \int_{\mathcal{X}} (1 + V(z))^2 [\tilde{\mu}_0(Y) + Q(Y_0)\tilde{\mu}_0(Y)](dz) \right] < \infty \]

by (22). A fortiori, \( E[\rho_1(\tilde{\mu}_0(Y_1), Q(Y_0)\tilde{\mu}_0(Y))] < \infty \), too. Now it follows from (35) and \( r_1(0) + r_2(0) < \infty \) that

\[ \sum_{n=1}^{\infty} E[\rho_1(\mu_n(Y), \mu_{n+1}(Y))] < \infty. \] (36)

Consequently, for a.e. \( \omega \in \Omega \), the sequence \( \mu_n(Y(\omega)), n \in \mathbb{N} \) is Cauchy and hence convergent for the metric \( \rho_1 \). Its limit is denoted by \( \mu_z(\omega) \).

For later use, we remark that \( \omega \rightarrow \int_{\mathcal{X}} \phi(z)\mu_z(\omega)(dz) \) is \( \sigma(Y) \)-measurable for every \( \phi \in \Phi(V) \). Hence, there is a measurable \( \Psi_\phi : \mathbb{Y} \rightarrow \mathbb{R} \) such that

\[ \Psi_\phi(Y) = \int_{\mathcal{X}} \phi(z)\mu_z(dz) \text{ a.s.} \] (37)

In the sequel, we will need the definition (28) for the kernel \((y, A) \rightarrow \mu_n(y)(A), \ y \in \mathbb{Y}, A \in \mathcal{B} \) and for similar kernels. Notice that, for any measurable function \( w : \mathcal{X} \rightarrow \mathbb{R}_+ \),

\[ \int_{\mathcal{X}} w(z) \left| \mathcal{E}[\mu_n(Y)] - \mathcal{E}[\mu_{n+1}(Y)] \right| (dz) \leq \int_{\mathcal{X}} w(z) \mathcal{E}\left[ |\mu_n(Y) - \mu_{n+1}(Y)| \right] (dz). \] (38)

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

\[ \int_{\mathcal{X}} w(z)\mathcal{E}\left[ |\mu_n(Y) - \mu_{n+1}(Y)| \right] (dz) = E \left[ \int_{\mathcal{X}} w(z) \left| \mu_n(Y) - \mu_{n+1}(Y) \right| (dz) \right]. \] (39)

Notice that \( \mu_n = \text{Law}(X_n) = \mathcal{E}[\mu_n(\tilde{Y}_n)] = \mathcal{E}[\mu_n(Y)] \). We infer from (38) and (39) that

\[ \rho_1(\mu_n, \mu_{n+1}) = \int_{\mathcal{X}} (1 + V(z)) \left| \mathcal{E}[\mu_n(Y)] - \mathcal{E}[\mu_{n+1}(Y)] \right| (dz) \]

\[ \leq \int_{\mathcal{X}} (1 + V(z))\mathcal{E}\left[ |\mu_n(Y) - \mu_{n+1}(Y)| \right] (dz) \]

\[ = E \left[ \int_{\mathcal{X}} (1 + V(z)) |\mu_n(Y) - \mu_{n+1}(Y)| (dz) \right] \]

\[ = E[\rho_1(\mu_n(Y), \mu_{n+1}(Y))]. \]
Then, it follows from (36) that
\[
\sum_{n=1}^{\infty} \rho_1(\mu_n, \mu_{n+1}) < \infty,
\]
so \( \mu_n, n \geq 0 \) is a Cauchy sequence for the complete metric \( \rho_1 \). Hence, it converges to some probability \( \mu_* \) as \( n \to \infty \). The claimed convergence rate also follows by the above estimates.

To show uniqueness, let \( X'_0 \) be another initial condition satisfying Assumption 2.7, with the corresponding \( \tilde{\mu}'_0(y) \), see (21). Defining, just like in (30),
\[
\mu'_0(y) := \tilde{\mu}'_0(y), \quad \mu'_n(y) := Q(y_0)Q(y_\cdot -1) \cdots Q(y_{n-1})\tilde{\mu}'_0(y_{n+1}), \quad n \geq 1,
\]
the estimates (34) and (35) show that
\[
\nu_1(\mathcal{E}[\mu'_n(Y)], \mathcal{E}[\mu_n(Y)]) \leq E[\nu_1(\mu'_n(Y), \mu_n(Y))] \\
\leq 2E[\nu_1(\tilde{\mu}_0(Y), \tilde{\mu}'_0(Y))] K(g(n))e^{-\frac{\eta}{2}(g(n))} \\
+ 8E^{1/2}\left[\nu_1^2(\tilde{\mu}_0(Y), \tilde{\mu}'_0(Y))\right] \sum_{k=n}^{\infty} \frac{K(g(k+1))}{\alpha(g(k+1))\zeta(g(k+1))}\sqrt{\ell(k)} \quad (40)
\]
which tends to 0 when \( n \to \infty \) since, as before, \( E[\nu_1^2(\tilde{\mu}_0(Y), \tilde{\mu}'_0(Y))] < \infty \), by Assumption 2.7.

**Remark 4.3** The proof of Theorem 2.12 also implies convergence for the “quenched” process: there is a set \( \mathcal{Y}' \subset \mathcal{Y} \) with \( \text{Law}(Y)\mathcal{Y}' = 1 \) such that, for all \( y \in \mathcal{Y}' \), the sequence \( \mu_n(y) \) converges in \( \rho_1 \) to a limiting probability \( \mu_z(y) \) as \( n \to \infty \).

**Remark 4.4** Define the probability \( \tilde{\mu}(A) := E[\mu_z(A)], A \in \mathcal{B} \). It is clear that, for every \( \phi \in \Phi(1) \),
\[
\int_{\mathcal{X}} \phi(z)\mu_z(dz) = \lim_{n \to \infty} \int_{\mathcal{X}} \phi(z)\mu_n(dz) = \lim_{n \to \infty} \int_{\mathcal{X}} \phi(z)\mathcal{E}[\mu_n(Y)](dz) \\
= \lim_{n \to \infty} E\left[\int_{\mathcal{X}} \phi(z)\mu_n(Y)(dz)\right] = E\left[\int_{\mathcal{X}} \phi(z)\mu_z(dz)\right] = \int_{\mathcal{X}} \phi(z)\tilde{\mu}(dz),
\]
hence \( \tilde{\mu} = \mu_* = E[\mu_z(Y)(A)] \), see the above remark.

**Proof of Theorem 2.16.** Estimates of Theorem 2.12 imply
\[
\rho_0(\mu_n(y_n), \mu_{n+1}(y_{n+1})) \leq (1 - \zeta(\tilde{y}_n)/4)^n \rho_1(\tilde{\mu}_0(\tilde{y}_{n+1}), Q(y_{n+1})\tilde{\mu}_0(\tilde{y}_{n})).
\]
By (29), this leads to
\[
\rho_0(\mu_n, \mu_{n+1}) \leq E[\rho_0(\mu_n(Y), \mu_{n+1}(Y))] \leq E[\rho_0(\mu_n(Y), \mu_{n+1}(Y))1_{\{M_n < g(n)\}}] + 2E[1_{\{M_n \geq g(n)\}}] \\
\leq (1 - \zeta(g(n))/4)^n E[\rho_1(\tilde{\mu}_0(\tilde{Y}_{-n+1}), Q(Y_{-n})\tilde{\mu}_0(\tilde{Y}_{-n}))1_{\{M_n < g(n)\}}] + 2P(M_n \geq g(n)) \\
\leq (1 - \zeta(g(n))/4)^n E[\rho_1(\tilde{\mu}_0(\tilde{Y}_1), Q(Y_0)\tilde{\mu}_0(Y))] + 2P(M_n \geq g(n)) \\
\leq C[e^{-n\zeta(g(n))/4} + \zeta(n)],
\]
for some \( C > 0 \), using Assumption 2.8 and (23). The result now follows as in the proof of Theorem 2.12.

**Remark 4.5** The convergence rates obtained by our method heavily depend on the choice of the functions \( g \) and \( \zeta \), for which there are multiple options. Hence, no optimality can be claimed at this level of generality. The approach, however, works in many concrete cases where available methods do not.

## 5 L-Mixing Processes

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

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

\[
\gamma_r(W, \tau) := \sup_{t \geq \tau} E^{1/r}[[W_t - E[W_t|G_{t-\tau}^+]]^r], \quad \tau \in \mathbb{N}, \quad \Gamma_r(W) := \sum_{\tau=0}^{\infty} \gamma_r(W, \tau).
\]

For some \( r \geq 1 \), the process \( W \) is called \textit{L-mixing of order} \( r \) with respect to \((G_t, G_t^+)\), \( t \in \mathbb{N} \) if it is adapted to \((G_t)_{t \in \mathbb{N}}\) and \( M_r(W) < \infty \), \( \Gamma_r(W) < \infty \). We say that \( W \) is \textit{L-mixing} if it is \( L \)-mixing of order \( r \) for all \( r \geq 1 \). This notion of mixing was introduced in [10].

**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, \( \Gamma_r(W) = \Gamma_r(W) \) and \( M_r(W) \leq 2M_r(W) \).

The next lemma is useful when checking the \( L \)-mixing property for a given process.

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

\[
E^{1/r}[|X - E[X|G]|^r] \leq 2E^{1/r}[|X - Y|^r].
\]
ProofSee Lemma 2.1 of [10].

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

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

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 [10].

$L$-mixing is, in many cases, easier to show than other, better-known mixing concepts such as $\alpha-$, $\beta-$ or $\phi$-mixing. There seems to be no implication between $L$-mixing and these latter conditions. For further information and related results, see [10].

6 Proofs of Ergodicity I

Throughout this section, let the assumptions of Theorem 2.17 be in force: $Y$ is an ergodic process; Assumptions 2.2 and 2.7 hold; Assumption 2.5 holds with $R(n) := 8K(n)/\lambda(n)$, $n \in \mathbb{N}$; and we have $r_1(0) + r_2(0) < \infty$ and

$$\left( \frac{K(g(N))}{\lambda(g(N))} \right)^{28} \frac{\pi(N)}{N} \to 0, \ N \to \infty.$$

In Sect. 4, we profited from contraction estimates for the metric $\rho_\beta$. These required, essentially, that given $X_t, X'_t$ a convenient coupling for $X_{t+1}, X'_{t+1}$ is realized, see, for example, (40). The exact nature of that coupling is hidden in Lemma 4.1. In the current section, we construct couplings for the whole process $X_t$ which allow us to show suitable mixing properties.

We now present a construction that is crucial for proving Theorem 2.17. 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 Sect. 5) for an auxiliary process ($Z$ below) which will, in turn, lead to Theorem 2.17. Such a representation with random mappings was used in [2, 11]. In our setting, however, there is also dependence on $y \in \mathcal{Y}$.

For $R \geq 0$, denote by $\mathcal{C}(R)$ 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

$$P(T_t(y, x, \cdot) \in A) = Q(y, x, A),$$

for all $t \geq 1$, $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $A \in \mathcal{B}$. There exist independent sigma-algebras $\mathcal{L}_t$, $t \geq 1$ such that the random variables $T_t(y, x, \cdot)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ are $\mathcal{L}_t$-measurable. There are events $J_t(y) \in \mathcal{L}_t$, for all $t \geq 1$, $y \in \mathcal{Y}$ such that

$$J_t(y) \subset \{ \omega \in \Omega : T_t(y, \cdot, \omega) \in \mathcal{C}(R(\|y\|)) \} \quad \text{and} \quad P(J_t(y)) \geq \alpha(\|y\|).$$
**Proof** Let $U_t, t \geq 1$ be an independent sequence of uniform random variables on $[0, 1]$. Let $\varepsilon_t, t \geq 1$ be another such sequence, independent of $(U_t)_{t \geq 1}$. By enlarging the probability space, if necessary, we can always construct such random variables and we may even assume that $(U_t, \varepsilon_t), t \geq 1$ are independent of $(X_0, (Y_t)_{t \in \mathbb{Z}})$. Let $L_t := \sigma(U_t, \varepsilon_t)$.

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 [8], 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 (42).

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 $\nu_n$) and define its $(\mathfrak{A} \otimes B(\mathbb{R})$-measurable) pseudoinverse by $j_n^-(y, z) := \inf \{r \in \mathbb{Q} : j_n(y, r) \geq z\}, z \in (0, 1)$. Here, $B(\mathbb{R})$ refers to the Borel field of $\mathbb{R}$. Similarly, for $y \in B_n$ and $x \in C(R(n))$, let

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

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

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

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

$$q^-(y, x, z) := \inf \{r \in \mathbb{Q} : q(y, x, r) \geq z\}, z \in (0, 1).$$

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

$$T_t(y, x, \omega) := q^-(y, x, \varepsilon_t), \text{ if } U_t(\omega) > \alpha(n) \text{ or } U_t(\omega) \leq \alpha(n) \text{ but } x \notin C(R(n)),
$$

$$T_t(y, x, \omega) := j_n^-(y, \varepsilon_t), \text{ if } U_t(\omega) \leq \alpha(n) \text{ and } x \in C(R(n)).$$

Notice that $T_t(y, \cdot, \omega) \in \mathcal{E}(R(n))$ whenever $U_t(\omega) \leq \alpha(n)$, this implies (42) with $J_t(y) := \{\omega \in \Omega : U_t(\omega) \leq \alpha(\|y\|)\}$. It is easy to check (41), 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 2.17.

We drop dependence of the mappings $T_t$ on $\omega \in \Omega$ in the notation from now on and will simply write $T_t(y, x)$. We continue our preparations for the proof of Theorem 2.17. Let

$$\mathcal{G}_0 := \{\emptyset, \Omega\}, \quad \mathcal{G}_t := \sigma(\varepsilon_i, U_i, 1 \leq i \leq t), t \geq 1 \tag{43}$$
and

$$G_t^+ := \sigma(\varepsilon_i, U_i, i \geq t + 1), \ t \in \mathbb{N}.$$  \hfill (44)

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, and it proceeds rather through establishing a convenient mixing property. The ensuing arguments will lead to Theorem 2.17 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 time-inhomogeneous. (And for us this is the crucial point.) See also Sect. 7 about these issues.

The main ideas of the arguments below go back to [2] and [11]. In [11], Doeblin chains were treated. We need to extend those arguments substantially in the present, more complicated setting.

Let us fix \( y = (y_j)_{j \in \mathbb{Z}} \in \mathcal{Y} \) till further notice such that, for some \( H \in \mathbb{N} \), \( \|y_j\| \leq H \) holds for all \( j \in \mathbb{Z} \). Define

$$Z_0 := X_0, \quad Z_{t+1} := T_{t+1}(y_t, Z_t), \ t \in \mathbb{N}.$$  \hfill (45)

Clearly, the process \( Z \) heavily depends on the choice of \( y \). However, for a while we do not signal this dependence for notational simplicity. Fix also \( m \in \mathbb{N} \) till further notice.

Define also

$$\tilde{Z}_m := \tilde{x}, \quad \tilde{Z}_{t+1} := T_{t+1}(y_t, \tilde{Z}_t), \ t \geq m.$$  \hfill (46)

Notice that \( \tilde{Z}_t, t \geq m \) are \( G_m^+ \)-measurable.

Our purpose will be to prove that, with a large probability, \( Z_{m+\tau} = \tilde{Z}_{m+\tau} \) for \( \tau \) large enough. In other words, a coupling between the processes \( Z \) and \( \tilde{Z} \) is realized. Fix \( \epsilon = \epsilon(H) > 0 \) 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 \( \overline{D} := \{ (x_1, x_2) \in \mathcal{X}^2 : V(x_1) + V(x_2) \leq R(H) \} \).

**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 implies that, for \( k \geq 1 \),

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

Assumption 2.7 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,

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

\[ \square \]

The counterpart of the above lemma for \( X \) 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 2.7. So, for each \( n \geq 1 \),

\[ E[V(X_n)] \leq \int_{\mathcal{X}} (1 + V(z))\mu_n(dz) \]

\[ \leq \int_{\mathcal{X}} (1 + V(z))|\mu_n - \mu_0|(dz) + \int_{\mathcal{X}} (1 + V(z))\mu_0(dz) \]

\[ = \rho_1(\mu_n, \mu_0) + E[V(X_0)] + 1 \]

\[ \leq \rho_1(\mu_n, \mu_*) + \rho_1(\mu_0, \mu_*) + E[V(X_0)] + 1 \]

\[ \leq 2C[r_1(0) + r_2(0)] + E[V(X_0)] + 1, \]

by Theorem 2.12. \[ \square \]

We note for later use that if \( z \in \mathcal{X} \setminus \mathcal{D} \), then for all \( y \in A_H \),

\[ [Q(y)(K(H) + V)](z) \leq (1 - \lambda(H))V(z) + 2K(H) \]

\[ \leq (1 - \lambda(H)/2)V(z). \]  

(47)

Recall Definitions (43) and (44). Define the \((G_t)_{t \in \mathbb{N}}\)-stopping times

\[ \sigma_0 := m, \quad \sigma_{n+1} := \min\{i > \sigma_n : \tilde{Z}_i \in \mathcal{D} \}. \]

The results below serve to control the number of returns to \( \mathcal{D} \) and the probability of coupling between the processes \( Z \) and \( \tilde{Z} \). Our estimation strategy in the proof of Theorem 2.17 will be the following. We will control \( P(\tilde{Z}_{\tau+m} \neq Z_{\tau+m}) \) for large \( \tau \): either there were only few returns of the process \( Z \) to \( \mathcal{D} \) (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 \( \mathcal{D} \).

**Lemma 6.5** There is \( \tilde{C} > 0 \) such that

\[ \sup_{n \geq 1} E \left[ \exp(\varphi(H)(\sigma_{n+1} - \sigma_n))|G_{\sigma_n} \right] \leq \frac{\tilde{C}}{\lambda^2(H)}. \]
and
\[ E[\exp(\varphi(H)(\sigma_1 - \sigma_0))] \leq \frac{\tilde{C}}{\lambda^2(H)} \]
where
\[ \varphi(H) := \ln(1 + \lambda(H)/2). \] (48)

In particular, \( \sigma_n < \infty \) a.s. for each \( n \in \mathbb{N} \). Furthermore, \( \tilde{C} \) does not depend on either \( y \), \( m \) or \( H \).

**Proof** We can estimate, for \( k \geq 1 \) and \( n \geq 1 \),

\[
P(\sigma_{n+1} - \sigma_n > k|G_{\sigma_n}) = P(\bar{Z}_{\sigma_n+k} \notin \bar{D}, \ldots, \bar{Z}_{\sigma_n+1} \notin \bar{D}|G_{\sigma_n})
\leq E \left[ \frac{V(Z_{\sigma_n+k}) + V(\tilde{Z}_{\sigma_n+k})}{R(H)} \right]_{\{Z_{\sigma_n+k-1} \notin \bar{D}\}} \cdots \{Z_{\sigma_n+1} \notin \bar{D}\}|G_{\sigma_n}]
= E \left[ \frac{V(Z_{\sigma_n+k}) + V(\tilde{Z}_{\sigma_n+k})}{R(H)} \right]_{\{Z_{\sigma_n+k-1} \notin \bar{D}\}} \{G_{\sigma_n+k-1}\} \cdots \{Z_{\sigma_n+1} \notin \bar{D}\}|G_{\sigma_n}].
\]

Notice that \( \{\bar{Z}_{\sigma_n+k-1} \notin \bar{D}\} \subset \{Z_{\sigma_n+k-1} \notin D\} \cup \{\tilde{Z}_{\sigma_n+k-1} \notin D\} =: E_1 \cup E_2 \). Assumption 2.2 and the observation (47) imply that

\[
E \left[ \frac{V(Z_{\sigma_n+k}) + V(\tilde{Z}_{\sigma_n+k})}{R(H)} \right]_{\{Z_{\sigma_n+k-1} \notin \bar{D}\}} \{G_{\sigma_n+k-1}\} \cdots \{Z_{\sigma_n+1} \notin \bar{D}\}|G_{\sigma_n}]
\leq 1_{E_1} \left[ \frac{1}{R(H)} [(1 - \lambda(H)/2)V(Z_{\sigma_n+k-1}) - K(H)] + \frac{1}{R(H)} [(1 - \lambda(H))V(\tilde{Z}_{\sigma_n+k-1}) + K(H)] \right]
+ 1_{E_2 \setminus E_1} \left[ \frac{1}{R(H)} [(1 - \lambda(H))V(Z_{\sigma_n+k-1}) - K(H)] + \frac{1}{R(H)} [(1 - \lambda(H)/2)V(\tilde{Z}_{\sigma_n+k-1}) + K(H)] \right]
\leq \frac{1 - \lambda(H)/2}{R(H)} [V(Z_{\sigma_n+k-1}) + V(\tilde{Z}_{\sigma_n+k-1})].
\]

This argument can clearly be iterated and leads to

\[
P(\sigma_{n+1} - \sigma_n > k|G_{\sigma_n}) \leq \frac{(1 - \lambda(H)/2)^{k-1}}{R(H)} \cdot E \left[ \frac{V(Z_{\sigma_n+1}) + V(\tilde{Z}_{\sigma_n+1})}{G_{\sigma_n}} \right]
\]
\[
\leq \frac{(1 - \lambda(H)/2)^{k-1}}{R(H)} \left[ (1 - \lambda(H)) \left[ V(Z_{\sigma_n}) + V(\tilde{Z}_{\sigma_n}) \right] + 2K(H) \right] \\
\leq (1 - \lambda(H)/2)^k,
\]

by Assumption 2.2, since \( Z_{\sigma_n} \in \mathcal{D} \). In the case \( n = 0, k \geq 1 \), we arrive at

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

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 \( n \geq 1 \),

\[
E \left[ \exp\{\varrho(H)(\sigma_{n+1} - \sigma_n)\} \right| G_{\sigma_n} \leq \sum_{k=0}^{\infty} e^{\varrho(H)(k+1)} \left( 1 - \frac{\lambda(H)}{2} \right)^k \\
\leq 2 \sum_{k=0}^{\infty} \left( 1 - \frac{\lambda^2(H)}{4} \right)^k = \frac{8}{\lambda^2(H)}.
\]

When \( n = 0 \), we obtain

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

for some \( \bar{C} \). We may and will assume \( \bar{C} \geq 8 \). The statement follows. \( \Box \)

The quantity \( \epsilon > 0 \) has been arbitrary thus far. Now, let us make the choice

\[
\epsilon := \epsilon(H) = \frac{\varrho(H)}{4(\ln(\bar{C}) - 2\ln(\lambda(H)))}.
\]

Recall that \( \tau \geq 1 \) has also been arbitrary. Recall that \( \vartheta = \lceil \epsilon \tau \rceil \).

**Corollary 6.6** If \( \tau \geq 1/\epsilon(H) \), 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 \( \tau \geq 1/\epsilon(H) \) implies

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

which guarantees

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

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

**Lemma 6.7** For all \( \tau \geq 1 \),

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

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

\[ P(Z_{m+\tau} \neq \tilde{Z}_{m+\tau}, \sigma(\vartheta) \leq m + \tau) \leq P(U_{\sigma(1) + 1} > \alpha(H), \ldots, U_{\sigma(\vartheta - 1) + 1} > \alpha(H)) \]

\[ = E[E[1_{U_{\sigma(\vartheta - 1) + 1} > \alpha(H)}|\mathcal{G}_{\sigma(\vartheta - 1)}]1_{U_{\sigma(1) + 1} > \alpha(H)} \cdots 1_{U_{\sigma(\vartheta - 2) + 1} > \alpha(H)}]]. \]

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 \( \phi \in \Phi(V^\delta) \) for some \( 0 < \delta \leq 1/2 \). 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, \( \Gamma_p(\phi(Z)), M_p(\phi(Z)) \) (see Sect. 5 for the definitions of these quantities) have upper bounds that do not depend on \( y \), only on \( H \).

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

\[ (x + y)^r \leq 2^{r-1}(x^r + y^r) \text{ if } r \geq 1; \quad (x + y)^r \leq x^r + y^r \text{ if } 0 < r < 1. \]
Proof of Lemma 6.8. Clearly,

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

with some constant \( \tilde{C} > 0 \), 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 \( \gamma_p(\phi(Z), \tau) \).

Since \( \tilde{Z}_m \) is deterministic, \( \tilde{Z}_{m+\tau} \) is \( \mathcal{G}_m^+ \)-measurable for \( \tau \geq 0 \). Lemma 5.2 implies that, for \( \tau \geq 1 \),

\[
E^{1/p}[(\phi(Z_{m+\tau}) - E[\phi(Z_{m+\tau})|\mathcal{G}_m^+])^p] \\
\leq 2E^{1/p}[(\phi(Z_{m+\tau}) - \phi(\tilde{Z}_{m+\tau}))^p] \\
\leq 2E^{1/p}[(|\phi(Z_{m+\tau})| + |\phi(\tilde{Z}_{m+\tau})|)^p1_{\{Z_{m+\tau} \neq \tilde{Z}_{m+\tau}\}}] \\
\leq 2\delta[(|\phi(Z_{m+\tau})| + |\phi(\tilde{Z}_{m+\tau})|)^{1/\delta}]P^{\frac{1-p\delta}{p}}(Z_{m+\tau} \neq \tilde{Z}_{m+\tau}), \tag{50}
\]

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

\[
E^\delta[(|\phi(Z_{m+\tau})| + |\phi(\tilde{Z}_{m+\tau})|)^{1/\delta}] \\
\leq \tilde{C} \left[ 1 + \left( E[V(X_0)] + \frac{K(H)}{\lambda(H)} \right)^{\delta} \right] \\
+ \tilde{C} \left[ 1 + \left( V(\tilde{x}) + \frac{K(H)}{\lambda(H)} \right)^{\delta} \right] \leq \tilde{C} \left[ \frac{K(H)}{\lambda(H)} \right]^{\delta}, \tag{51}
\]

for some suitable \( \tilde{C} > 0 \). Here, we have used \( K \geq 1 \). 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 (51), Lemma 6.7 and Corollary 6.6 that for \( \tau \geq 1/\varepsilon(H) \),

\[
\gamma_p(\phi(Z), \tau) \\
\leq 2\tilde{C} \left( \frac{K(H)}{\lambda(H)} \right)^{\delta} P^{\frac{1-p\delta}{p}}(Z_{m+\tau} \neq \tilde{Z}_{m+\tau}) \\
\leq 2\tilde{C} \left( \frac{K(H)}{\lambda(H)} \right)^{\delta} \left[ \exp(-\alpha(H)[\varepsilon(H)\tau - 1]) + \exp\left( -\frac{\varrho(H)\tau}{2} \right) \right]^{(1-p\delta)/p}, \tag{52}
\]

\( \tilde{C} \) Springer
noting that the estimates of Lemma 6.7 and Corollary 6.6 do not depend on the choice of \( m \). For each integer 

\[ 0 \leq \tau < \frac{1}{\epsilon(H)}, \]

we will apply the trivial estimate

\[ \gamma_p(\phi(Z), \tau) \leq 2M_p(\phi(Z)) \leq 2M_1(\phi(Z)) \leq 2\hat{C} \left[ \frac{K(H)}{\lambda(H)} \right]^\delta, \quad (53) \]

recall (51). Keep in mind the definition \( \vartheta = \lceil \epsilon \tau \rceil \). We can then write, using the formula for the sum of geometric series,

\[ \Gamma_p(\phi(Z)) \leq 2\hat{C} \left[ \frac{K(H)}{\lambda(H)} \right]^\delta \]

\[ \left[ \frac{1}{\epsilon(H)} + \sum_{\tau \geq \frac{1}{\epsilon(H)}} \left[ \exp(-\alpha(H)[\epsilon(H)\tau - 1]) + \exp\left( -\frac{\vartheta(H)\tau}{2} \right) \right]^{(1-p\delta)/p} \right] \]

\[ \leq c' \left[ \frac{1}{\epsilon(H)} + \exp\left( \frac{\alpha(H)(1-p\delta)/p}{1-\exp\left( -\frac{\vartheta(H)(1-p\delta)}{2p} \right)} \right) \right] \]

\[ \left( \frac{K(H)}{\lambda(H)} \right)^\delta \quad (54) \]

with some constant \( c' \). Using elementary properties of the functions \( x \to 1/(1-e^{-x}) \) and \( x \to \ln(1+x) \) and the definitions (49) and (48), \( \Gamma_p(\phi(Z)) \) can be estimated from above by

\[ c'' \left[ \frac{1}{\alpha(H)\epsilon(H)} + \frac{1}{\lambda(H)} \right] \left( \frac{K(H)}{\lambda(H)} \right)^\delta \leq c''' \frac{\ln(\lambda(H))}{\alpha(H)\lambda(H)} \left( \frac{K(H)}{\lambda(H)} \right)^\delta \quad (55) \]

with some \( c'', c''' > 0 \). The \( L \)-mixing property of order \( p \) follows. (Note, however, that \( c''' \) depends on \( p, \delta \) as well as on \( V(\tilde{x}) \) and \( E[V(X_0)] \).) \( \square \)

**Proof of Theorem 2.17.** Recall the definitions (45) and (46). Now, we start signalling the dependence of \( Z \) on \( y \) and hence write \( Z_t^y, t \in \mathbb{N} \). Note that the law of \( Z_t^\tilde{y}, t \in \mathbb{N} \) equals that of \( X_t, t \in \mathbb{N} \), by construction of \( Z \) and by Remark 6.2.

For \( t \in \mathbb{N} \) and \( y \in \mathcal{Y} \), define \( \psi_t(y) := \int_X \phi(x)\mu_n(y)(dx) \), recall the definition of \( \mu_n(y) \) from (30). Notice that \( \psi_t(\tilde{y}_t) = E[\phi(Z_t^\tilde{y})] \). Define \( W_t(y) := \phi(Z_t^y) - \psi_t(\tilde{y}_t) \). Clearly, \( W_t(y) \) is a zero-mean process.

Fix \( p \geq 2 \) such that \( \delta p < 1 \). Fix \( N \in \mathbb{N} \) for the moment. In the particular case where \( y \) satisfies \( |y_j| \leq H := g(N) \), \( j \in \mathbb{N} \), the process \( W_t(y), t \in \mathbb{N} \) is \( L \)-mixing by
Lemma 6.8 and Remark 5.1. Hence, Lemma 5.3 implies

\[ E^{1/p} \left[ \frac{W_1(y) + \ldots + W_N(y)}{N} \right] \]

\[ \leq C_p M_p^{1/2} (W(y)) \Gamma_p^{1/2} (W(y)) \]

\[ \leq \frac{C_p M_1^{1/\delta} (W(y)) \Gamma_p^{1/2} (W(y))}{N^{1/2}} \]

\[ \leq \frac{2C_p \sqrt{C} [K(g(N))/\lambda(g(N))]^{\delta/2} \sqrt{(K(g(N))/\lambda(g(N)))^{\delta/2} \pi^{1/2} (N)}}{N^{1/2}} \]

by (51) and (55); recall also Remark 5.1. Fix \( \tilde{y} \in A_0 \) and define

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

Let \( \tilde{Y} = (\tilde{Y}_j)_{j \in \mathbb{Z}} \in \mathfrak{Y} \). Note that, by \( \phi \in \Phi(V^\delta) \),

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

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

\[ \frac{E^{1/p} \left[ \frac{W_1(Y) + \ldots + W_N(Y)}{N} \right]^p}{E^{1/p} \left[ \frac{W_1(\tilde{Y}) + \ldots + W_N(\tilde{Y})}{N} \right]^p} \]

\[ + M_{1/\delta} (W(Y)) P^{1-pq} \left( (\tilde{Y}_1, \ldots, \tilde{Y}_N) \neq (Y_1, \ldots, Y_N) \right) \]

\[ \leq \frac{C' (K(g(N))/\lambda(g(N)))^{\delta/2} \pi^{1/2} (N)}{N^{1/2}} + C' \left( 1 + \sup_{n \in \mathbb{N}} E[V(X_n)] \right)^{\delta} \ell^{1-pq} (N) \]

\[ \leq \frac{C'' (K(g(N))/\lambda(g(N)))^{\delta/2} \pi^{1/2} (N)}{N^{1/2}} + C'' \ell^{1-pq} (N), \quad (56) \]

with some constants \( C', C'' > 0 \), by Lemma 6.4. Here, we have also used the fact that if \((\tilde{Y}_1, \ldots, \tilde{Y}_N) = (Y_1, \ldots, Y_N)\), then also \( W_j(Y) = W_j(\tilde{\tilde{Y}}) \), \( j = 1, \ldots, N \). The quantity (56) tends to 0 by our hypotheses (12) and \( r_2(0) < \infty \).

Recall the notation \( \hat{Y}_n := (Y_{j+n})_{j \in \mathbb{Z}} \) and the functional \( \Psi_\phi \) from (37). Now, we can estimate

\[ \left| \int_X \phi(z) \mu_\phi (dz) - \frac{\sum_{j=1}^N \phi(Z_{j})}{N} \right| \]

\[ \leq \left| \int_X \phi(z) \mu_\phi (dz) - \frac{\sum_{j=1}^N \Psi_\phi (\tilde{Y}_j)}{N} \right| \]
\[ + \left| \sum_{j=1}^{N} \frac{\Psi(\hat{Y}_j)}{N} - \sum_{j=1}^{N} \frac{\Psi(\hat{Y}_j)}{N} \right| \]
\[ + \left| \sum_{j=1}^{N} \frac{\psi_j(\hat{Y}_j)}{N} - \sum_{j=1}^{N} \phi(Z_j) \right|. \tag{57} \]

Birkhoff’s theorem and the ergodicity of the process \( Y \) imply that
\[ \sum_{j=1}^{N} \frac{\Psi(\hat{Y}_j)}{N} \to \int_{\mathcal{X}} \phi(z) \mu_\ast(dz), \ N \to \infty, \]
almost surely, hence also in probability, noting Remark 4.4.

By stationarity of the process \( \hat{Y}_k, k \in \mathbb{Z} \), we have that
\[ E \left[ \max \left( |\psi_j(\hat{Y}_j) - \Psi_{\phi}(\hat{Y}_j)|, 1 \right) \right] = E \left[ \max \left( |\psi_j(Y) - \Psi_{\phi}(Y)|, 1 \right) \right]. \]

The proof of Theorem 2.12 (see the discussion after (36)) and Remark 4.4 shows that
\[ \psi_j(Y) \to \Psi_{\phi}(Y) \] almost surely as \( j \to \infty \). It follows that \( \psi_j(\hat{Y}_j) - \Psi_{\phi}(\hat{Y}_j) \) tends to 0 in probability and so does the second term on the right-hand side of (57).

The third term on the right-hand side of (57) equals
\[ \left| \frac{W_1(Y) + \ldots + W_N(Y)}{N} \right|. \]

We claim that it converges to 0 in probability. Notice that \( \ell(N) \to 0, N \to \infty \) by the hypothesis \( r_2(0) < \infty \). It was a hypothesis of Theorem 2.17 that \( \pi(N)(K(N)/\lambda(N))^{2\beta}/N \) tends to 0 and hence, actually, (56) implies that the term in question tends to 0 in \( L^p \), a fortiori, in probability.

To sum up,
\[ \left| \int_{\mathcal{X}} \phi(z) \mu_\ast(dz) - \frac{\sum_{j=1}^{N} \phi(X_j)}{N} \right| \to 0 \]
in probability, recalling that the laws of \( Z_n^Y, n \in \mathbb{N} \) and \( X_n, n \in \mathbb{N} \) coincide.

To show convergence in \( L^p \), it suffices to check the uniform integrability of the family of random variables \( V^{2\beta}(X_n), n \in \mathbb{N} \) since \( \phi \in \Phi(V^{2\beta}) \). This follows from \( p < 1/\delta \) and from Lemma 6.4. The theorem has been shown for \( p \geq 2 \) but this implies the result for \( 1 \leq p < 2 \), too. \( \square \)

**Remark 6.9** In the proof of Theorem 2.17, we could find estimates for the \( L^p \) convergence rate for the third term in (57) (see (56)). We would be able to do likewise for the second term, using the estimates in our arguments. However, there is, a priori, no
rate estimate for
\[ h_N := \left| \frac{1}{N} \sum_{j=1}^{N} \Psi_{\phi}(\hat{Y}_j) - \int_{\mathcal{X}} \phi(z) \mu_*(dz) \right|, \]

as this depends on the mixing properties of \( Y \). Making suitably strong assumptions about the process \( Y \), however, this term could also be estimated. In the ideal case, \( E^{1/p}[h_N^p] \) is of the order \( 1/\sqrt{N} \).

7 Ramifications

Let \( X_t, \ t \in \mathbb{N} \) be a \( \mathcal{X} \)-valued time-inhomogeneous Markov chain. Denote by \( Q_t(x, A) \) the transition kernel at times \( t \geq 1 \). We impose that there exist \( \lambda, K > 0 \) such that, for all \( t \geq 1 \),
\[ [Q_t V](x) \leq (1 - \lambda) V(x) + K, \ x \in \mathcal{X}, \]
for some measurable function \( V : \mathcal{X} \to \mathbb{R}_+ \). Furthermore, for some \( \alpha > 0 \) and for all \( t \geq 1 \),
\[ \inf_{x \in \mathcal{C}} Q_t(x, A) \geq \alpha \nu(A), \ A \in \mathcal{B} \]
for some probability \( \nu \) and for
\[ C := \{ x \in \mathcal{X} : V(x) \leq 4K/\lambda \}. \]

A simplified form of the argument of Lemma 6.1 gives us independent (but not identically distributed) random mappings \( T_t : \mathcal{X} \times \Omega \to \mathcal{X} \) such that \( P(T_t(x, \cdot) \in A) = Q_t(x, A) \) for \( t \geq 1 \). Note that \( T_t \) are only independent, but not identically distributed this time. Define \( Z_0 := X_0, Z_{t+1} := T_{t+1}(Z_t), t \in \mathbb{N} \), where we dropped, as before, the dependence of \( T_{t+1} \) on \( \omega \in \Omega \) in the notation.

Fix \( m \in \mathbb{N} \) and define \( \tilde{Z}_m := \tilde{x}, \tilde{Z}_{t+1} := T_{t+1}(\tilde{Z}_t), \ t \geq m \), for some fixed \( \tilde{x} \in \mathcal{X} \). Repeating the arguments of Sect. 6, we get the following result.

**Theorem 7.1** For suitable constants \( c_1, c_2 > 0 \),
\[ P(Z_m + \tau \neq \tilde{Z}_m + \tau) \leq c_1 \exp \left( -c_2 \frac{\alpha \lambda}{|\ln(\lambda)|} \tau \right), \ \tau \geq 1. \]

Furthermore, for \( 0 < \delta \leq 1/2 \) and for any \( \phi \in \Phi(V^\delta) \), \( \phi(X_t), t \in \mathbb{N} \) is \( L \)-mixing of order \( p \) for each \( 1 \leq p < 1/\delta \) and the following estimates hold:
\[ M_p(\phi(X)) = O((K/\lambda)^\delta) \text{ and } \Gamma_p(\phi(X)) = O \left( \left( \frac{K}{\lambda} \right)^\delta \frac{|\ln(\lambda)|}{\alpha \lambda} \right). \]
Although this result forms a very particular case of our framework, it is new and of considerable interest: it establishes a useful mixing property for functionals of a wide class of (even inhomogeneous) Markov processes.

8 Proofs of Ergodicity II

Proof of Theorem 2.18. This follows very closely the proof of Theorem 2.17, and 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 $\infty$ in the estimates (50). This leads to

$$\Gamma_p(\phi(Z)) \leq c''' \left| \ln(\lambda(H)) \right| = c''' \pi(H),$$

using the argument of (55). Then, the proof of convergence in probability can be completed as above. Note that instead of

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

we may write

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

in (56) and that

$$E^{1/p} \left[ \frac{W_1(y) + \ldots + W_N(y)}{N} \right]^p \leq \frac{C_p S^{1/2} c''' \pi(N)}{N^{1/2}}$$

in this case. As $\phi$ is bounded, $L^p$ convergence for all $p \geq 1$ also follows. $\square$

Acknowledgements 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 and the second author by the “Lendület” grant LP 2015-6 of the Hungarian Academy of Sciences and by the NKFIH grant K 143529. We thank Attila Lovas for pointing out two mistakes and for suggesting improvements. We also thank three anonymous referees for helpful comments.

Author Contributions The authors have no relevant financial or non-financial interests to disclose. Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Funding Open access funding provided by ELKH Alfréd Rényi Institute of Mathematics.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

1. Bhattacharya, R.N., Waymire, E.C.: Stochastic Processes with Applications. Wiley, New York (1990)
2. Bhattacharya, R., Waymire, E.C.: An approach to the existence of unique invariant probabilities for Markov processes. In: Limit Theorems in Probability and Statistics, János Bolyai Math. Soc., I (Baloghonye 1999), pp. 181–200 (2002)
3. Borovkov, A.A.: Ergodicity and Stability of Stochastic Processes. Wiley, New York (1998)
4. Cogburn, R.: The ergodic theory of Markov chains in random environments. Z. Wahrsch. Verw. Gebiete 66, 109–128 (1984)
5. Cogburn, R.: On direct convergence and periodicity for transition probabilities of Markov chains in random environments. Ann. Probab. 18, 642–654 (1990)
6. Comte, F., Renault, É.: Long memory in continuous-time stochastic volatility models. Math. Finance 8, 291–323 (1998)
7. Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. Quant. Finance 1, 223–236 (2001)
8. Dellacherie, C., Meyer, P.-A.: Probability and Potential. North-Holland, Amsterdam (1979)
9. Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is rough. Quantitative Finance 18, 933–949 (2018)
10. Gerencsér, L.: On a class of mixing processes. Stochastics 26, 165–191 (1989)
11. Gerencsér, L., Molnár-Sáska, G., Michaletzky, Gy., Tusnády, G., Vágó, Zs.: 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, pp. 2272–2277, IEEE Press, New York (2002)
12. Hairer, M., Mattingly, J.: 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.), Springer, Basel, (2011)
13. Kifer, Y.: Perron-Frobenius theorem, large deviations, and random perturbations in random environments. Math. Zeitschrift 222, 677–698 (1996)
14. Kifer, Y.: Limit theorems for random transformations and processes in random environments. Trans. Am. Math. Soc. 350, 1481–1518 (1998)
15. Lovas, A., Rásonyi, M.: Markov chains in random environment with applications in queueing theory and machine learning. Stochastic Process. Appl. 137, 294–326 (2021)
16. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Springer (1993)
17. Orey, S.: Markov chains with stochastically stationary transition probabilities. Ann. Probab. 19, 907–928 (1991)
18. Seppäläinen, T.: Large deviations for Markov chains with random transitions. Ann. Probab. 22, 713–748 (1994)
19. Stenflo, Ö.: Markov chains in random environments and random iterated function systems. Trans. Am. Math. Soc. 353, 3547–3562 (2001)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.