Commun. Math. Phys. 278, 253–288 (2008)
Digital Object Identifier (DOI) 10.1007/s00220-007-0390-4

Communications in
**Mathematical
Physics**

Lingering Random Walks in Random Environment on a Strip

Erwin Bolthausen[1], **Ilya Goldsheid**[2]

[1] Universität Zürich, Institut für Mathematik, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: eb@math.unizh.ch
[2] School of Mathematical Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, UK. E-mail: I.Goldsheid@qmul.ac.uk

Received: 30 July 2007 / Accepted: 1 November 2007
Published online: 8 December 2007 –

Abstract: We consider a recurrent random walk (RW) in random environment (RE) on a strip. We prove that if the RE is i. i. d. and its distribution is not supported by an algebraic subsurface in the space of parameters defining the RE then the RW exhibits the $(\log t)^2$ asymptotic behaviour. The exceptional algebraic subsurface is described by an explicit system of algebraic equations.

One-dimensional walks with bounded jumps in a RE are treated as a particular case of the strip model. If the one dimensional RE is i. i. d., then our approach leads to a complete and constructive classification of possible types of asymptotic behaviour of recurrent random walks. Namely, the RW exhibits the $(\log t)^2$ asymptotic behaviour if the distribution of the RE is not supported by a hyperplane in the space of parameters which shall be explicitly described. And if the support of the RE belongs to this hyperplane then the corresponding RW is a martingale and its asymptotic behaviour is governed by the Central Limit Theorem.

1. Introduction

The aim of this work is to describe conditions under which a recurrent random walk in a random environment (RWRE) on a strip exhibits the $\log^2 t$ asymptotic behaviour. This slow, lingering movement of a walk was discovered by Sinai in 1982 [18]. At the time, this work had brought to a logical conclusion the study of the so called simple RWs (SRW) started by Solomon in [19] and by Kesten, Kozlov, and Spitzer in [14]. The somewhat misleading term "simple" is often used as an abbreviation describing a walk on a one-dimensional lattice with jumps to nearest neighbours.

Our work was motivated by a question asked by Sinai in [18] about the validity of his (and related) results for other models. Perhaps the simplest extension of the SRW is presented by a class of one-dimensional walks whose jumps (say) to the left are bounded and to the right are of length at most one. These models were successfully studied by a number of authors and the relevant references can be found in [2]. We would like to quote one result concerning this special case since it is perhaps most close to our results stated below in Theorems 2 and 3. Namely, Bremont proved in [3] that if the environment is defined by a Gibbs measure on a sub-shift of finite type, then the asymptotic behaviour of a recurrent RW is either as in Sinai's theorem, or it is governed by the Central Limit Law.

General 1DWBJ were also studied by different authors. Key in [15] found conditions for recurrence of a wide class of 1DWBJ. Certain sufficient conditions for the Sinai behaviour of 1DWBJ were obtained by Letchikov in [17]. The results from [17] will be discussed in a more detailed way in Sect. 1.1 after the precise definition of the one-dimensional model is given. We refer the reader to [20] for further historical comments as well as for a review of other recent developments.

The main object of this paper is the RWRE on a strip. We prove (and this is the main result of this paper) that recurrent walks in independent identically distributed (i. i. d.) random environments on a strip exhibit the $\log^2 t$ asymptotic behaviour if the support of the distribution of the parameters defining the random environment does not belong to a certain algebraic subsurface in the space of parameters. This subsurface is defined by an explicit system of algebraic equations.

The one dimensional RW with bounded jumps can be viewed as a particular case of a RWRE on a strip. This fact was explained in [1] and we shall repeat this explanation here. Due to this reduction, our main result implies a complete classification of recurrent 1DWBJ in i.i.d. environments. Namely, the corresponding system of algebraic equations reduces in this case to one linear equation which defines a hyperplane in the space of parameters. If the support of the distribution of parameters does not belong to this hyperplane, then the RW exhibits the Sinai behaviour (see Theorem 2 below). But if it does, then (Theorem 3 below) the corresponding random walk is a martingale and its asymptotic behaviour is governed by the Central Limit Law. In brief, recurrent 1DWBJ are either of the Sinai type, or they are martingales.

In the case of a strip, a complete classification can also be obtained and it turns out that once again the asymptotic behaviour is either the Sinai, or is governed by the Invariance Principle. However, this case is less transparent and more technical even to describe in exact terms and we shall leave it for a future work.

The paper is organized as follows. We state Sinai's result and define a more general one-dimensional model in Sect. 1.1. Section 1.2 contains the definition of the strip model and the explanation of the reduction of the one-dimensional model to the strip case. Main results are stated in Sect. 1.3. Section 2 contains several statements which are then used in the proof of the main result, Theorem 1. In particular, we introduce random transformations associated with random environments in Sect. 2.2. It turns out to be natural to recall and to extend slightly, in the same Sect. 2.2, those results from [1] which are used in this paper. An important Lemma 5 is proved in Sect. 2.3; this lemma allows us to present the main algebraic statement of this work in a constructive form. In Sect. 2.4 we prove the invariance principle for the log of a norm of a product of certain matrices. This function plays the role of the so-called potential of the environment and is responsible for the Sinai behaviour of the random walk. It is used in the proof of our main result in Sect. 3.

Finally the Appendix contains results of which many (if not all) are not new but it is convenient to have them in a form directly suited for our purposes. Among these, the most important for our applications is the Invariance Principle (IP) for "contracting" Markov chains (Sect. 4.1.3). Its proof is derived from a well known IP for general Markov chains which, in turn, is based on the IP for martingales.

Conventions. The following notations and terminology shall be used throughout the paper. $\mathbb{R}$ is the set of real numbers, $\mathbb{Z}$ is the set of integer numbers, and $\mathbb{N}$ is the set of positive integers.

For a vector $x = (x_i)$ and a matrix $A = (a(i, j))$ we put

$$\|x\| \stackrel{\text{def}}{=} \max_i |x_i|, \quad \|A\| \stackrel{\text{def}}{=} \max_i \sum_j |a(i, j)|.$$

Note that $\|A\| = \sup_{|x|=1} \|Ax\|$. We say that A is strictly positive (and write $A > 0$), if all its matrix elements satisfy $a(i, j) > 0$. A is called non-negative (and we write $A \geq 0$), if all $a(i, j)$ are non negative. A similar convention applies to vectors.

1.1. Sinai's result and some of its extensions to 1DWBJ. Let $\omega \stackrel{\text{def}}{=} (p_n)_{-\infty<n<\infty}$ be a sequence of independent identically distributed (i. i. d.) random variables, satisfying $\varepsilon \leq p_n \leq 1 - \varepsilon$, where $\varepsilon > 0$. Put $q_n = 1 - p_n$ and consider a random walk $\xi(t)$ on a one-dimensional lattice with a starting point $\xi(0) = 0$ and transition probabilities

$$Pr_\omega\{\,\xi(t+1) = n+1 \,|\, \xi(t) = n\,\} = p_n, \quad Pr_\omega\{\,\xi(t+1) = n-1 \,|\, \xi(t) = n\,\} = q_n,$$

thus defining a measure $Pr_\omega\{\cdot\}$ on the space of trajectories of the walk. It is well known (Solomon, [19]) that this RW is recurrent in almost all environments ω if and only if $\mathbb{E} \ln \frac{q_n}{p_n} = 0$ (here $\mathbb{E}$ denotes the expectation with respect to the relevant measure $\mathbb{P}$ on the space of sequences). In [18] Sinai proved that if $\mathbb{E}(\ln \frac{q_n}{p_n})^2 > 0$ and $\xi(\cdot)$ is recurrent then there is a weakly converging sequence of random variables $b_t(\omega),\ t = 1, 2, \ldots$ such that

$$(\log t)^{-2}\xi(t) - b_t \to 0 \ \text{ as } \ t \to \infty. \tag{1.1}$$

The convergence in (1.1) is in probability with respect to the so-called annealed probability measure $\mathbb{P}(d\omega)Pr_\omega$ (for precise statements see Sect. 1.3). The limiting distribution of b_t was later found, independently, by Golosov [7,8] and Kesten [13].

The one-dimensional walk with bounded jumps on $\mathbb{Z}$ is defined similarly to the simple RW. Namely let $\omega \stackrel{\text{def}}{=} (p(n, \cdot)),\ n \in \mathbb{Z}$, be a sequence of non-negative vectors with $\sum_{k=-m}^{m} p(n, k) = 1$ and $m > 1$. Put $\xi(0) = 0$ and

$$Pr_\omega\,(\xi(t+1) = n + k \,|\, \xi(t) = n) \stackrel{\text{def}}{=} p(n, k), \quad n \in \mathbb{Z}. \tag{1.2}$$

Suppose next that $p(n, \cdot)$ is a random stationary in n (in particular it can be i. i. d.) sequence of vectors. Sinai's question can be put as follows: given that a RW is recurrent, what kind of asymptotic behaviour would one observe, and under what conditions?

There were several attempts to extend Sinai's result to the (1.2) model. In particular, Letchikov [17] proved that if for some $\varepsilon > 0$ with $\mathbb{P}$-probability 1

$$p(n, 1) \geq \sum_{k=-m}^{-2} p(n, k) + \varepsilon \ \text{ and } \ p(n, -1) \geq \sum_{k=2}^{m} p(n, k) + \varepsilon$$

and the distribution of the i. i. d. random vectors $p(n, \cdot)$ is absolutely continuous with respect to the Lebesgue measure (on the relevant simplex), then the analogue of Sinai's theorem holds. (In [17], there are also other restrictions on the distribution of the RE but they are much less important than the ones listed above.)

The technique we use in this work is completely different from that used in [2,3,15,17]. It is based on the methods from [1] and [6] and this work presents further development of the approach to the analysis of the RWRE on a strip started there.

1.2. Definition of the strip model. The description of the strip model presented here is the same as in [1].

Let (P_n, Q_n, R_n), $-\infty < n < \infty$, be a strictly stationary ergodic sequence of triples of $m \times m$ matrices with non-negative elements such that for all $n \in \mathbb{Z}$ the sum $P_n + Q_n + R_n$ is a stochastic matrix,

$$(P_n + Q_n + R_n)\mathbf{1} = \mathbf{1}, \tag{1.3}$$

where $\mathbf{1}$ is a column vector whose components are all equal to 1. We write the components of P_n as $P_n(i, j)$, $1 \le i, j \le m$, and similarly for Q_n and R_n. Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{T})$ be the corresponding dynamical system with Ω denoting the space of all sequences $\omega = (\omega_n) = ((P_n, Q_n, R_n))$ of triples described above, $\mathcal{F}$ being the corresponding natural σ-algebra, $\mathbb{P}$ denoting the probability measure on $(\Omega, \mathcal{F})$, and $\mathcal{T}$ being a shift operator on Ω defined by $(\mathcal{T}\omega)_n = \omega_{n+1}$. For fixed ω we define a random walk $\xi(t)$, $t \in \mathbb{N}$ on the strip $\mathbb{S} = \mathbb{Z} \times \{1, \ldots, m\}$ by its transition probabilities $\mathcal{Q}_\omega(z, z_1)$ given by

$$\mathcal{Q}_\omega(z, z_1) \overset{\text{def}}{=} \begin{cases} P_n(i, j) & \text{if } z = (n, i),\ z_1 = (n+1, j), \\ R_n(i, j) & \text{if } z = (n, i),\ z_1 = (n, j), \\ Q_n(i, j) & \text{if } z = (n, i),\ z_1 = (n-1, j), \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

This defines, for any starting point $z = (n, i) \in \mathbb{S}$ and any ω, a law $Pr_{\omega,z}$ for the Markov chain $\xi(\cdot)$ by

$$Pr_{\omega,z}(\xi(1) = z_1, \ldots, \xi(t) = z_t) \overset{\text{def}}{=} \mathcal{Q}_\omega(z, z_1)\mathcal{Q}_\omega(z_1, z_2) \cdots \mathcal{Q}_\omega(z_{t-1}, z_t). \tag{1.5}$$

We call ω the *environment* or the *random environment* on a strip $\mathbb{S}$. Denote by Ξ_z the set of trajectories $\xi(\cdot)$ starting at z. $Pr_{\omega,z}$ is the so-called quenched probability measure on Ξ_z. The semi-direct product $\mathbb{P}(d\omega)Pr_{\omega,z}(d\xi)$ of $\mathbb{P}$ and $Pr_{\omega,z}$ is defined on the direct product $\Omega \times \Xi_z$ and is called the annealed measure. All our main results do not depend on the choice of the starting point z. We therefore write Pr_ω instead of $Pr_{\omega,z}$ when there is no danger of confusion.

The one-dimensional model (1.2) reduces to a RW on a strip due to the following geometric construction. Note first that it is natural to assume (and we shall do so) that at least one of the following inequalities holds:

$$\mathbb{P}\{\omega : p(x, m) > 0\} > 0 \ \text{ or } \ \mathbb{P}\{\omega : p(x, -m) > 0\} > 0. \tag{1.6}$$

Consider the one-dimensional lattice as a subset of the X-axis in a two-dimensional plane. Cut this axis into equal intervals of length m so that each of them contains exactly m consecutive integer points. Turn each of these intervals around its left most integer point anti-clockwise by $\pi/2$. The image of $\mathbb{Z}$ obtained in this way is a part of a strip with distances between layers equal to m. Re-scaling the X-axis of the plane by m^{-1} makes the distance between the layers equal to one. The random walk on the line is thus transformed into a random walk on a strip with jumps to nearest layers.

The formulae for matrix elements of the corresponding matrices P_n, Q_n, R_n result now from a formal description of this construction. Namely, present $x \in \mathbb{Z}$ as $x = nm+i$, where $1 \le i \le m$. This defines a bijection $x \leftrightarrow (n, i)$ between the one-dimensional lattice $\mathbb{Z}$ and the strip $\mathbb{S} = \mathbb{Z} \times \{1, \ldots, m\}$. This bijection naturally transforms the

ξ-process on $\mathbb{Z}$ into a walk on $\mathbb{Z} \times \{1, \ldots, m\}$. The latter is clearly a random walk of type (1.5) and the corresponding matrix elements are given by

$$\begin{aligned} P_n(i, j) &= p(nm + i, m + j - i), \\ R_n(i, j) &= p(nm + i, j - i), \\ Q_n(i, j) &= p(nm + i, -m + j - i). \end{aligned} \tag{1.7}$$

1.3. Main results . Denote by $\mathcal{J}$ the following set of triples of $m \times m$ matrices:

$$\mathcal{J} \stackrel{\text{def}}{=} \{(P, Q, R) \,:\, P \geq 0,\ Q \geq 0,\ R \geq 0 \ \text{ and } \ (P + Q + R)\mathbf{1} = \mathbf{1}\}.$$

Let $\mathcal{J}_0 \subset \mathcal{J}$ be the support of the probability distribution of the random triple (P_n, Q_n, R_n) defined above (obviously, this support does not depend on n). The two assumptions **C1** and **C2** listed below will be referred to as Condition **C**.

Condition **C**

C1 (P_n, Q_n, R_n), $-\infty < n < \infty$, *is a sequence of independent identically distributed random variables.*

C2 *There is an* $\varepsilon > 0$ *and a positive integer number* $l < \infty$ *such that for any* $(P, Q, R) \in \mathcal{J}_0$ *and all* $i,\ j \in [1, m]$,

$$||R^l|| \leq 1 - \varepsilon, \ \ ((I - R)^{-1}P)(i, j) \geq \varepsilon, \ \ ((I - R)^{-1}Q)(i, j) \geq \varepsilon.$$

Remarks. 1. We note that say $((I - R_n)^{-1}P_n)(i, j)$ is the probability for a RW starting from (n, i) to reach $(n+1, j)$ at its first exit from layer n. The inequality $||R_n^l|| \leq 1-\varepsilon$ is satisfied in essentially all interesting cases and, roughly speaking, means that the probability for a random walk to remain in layer n after a certain time l is small uniformly with respect to n and ω.

2. If the strip model is obtained from the one-dimensional model, then **C2** may not be satisfied by matrices (1.7). This difficulty can be overcome if we replace **C2** by a much milder condition, namely:

C3 *For* $\mathbb{P}$ - *almost all* ω:

(a) *the strip* $\mathbb{S}$ *is the* (*only*) *communication class of the walk*,

(b) *there is an* $\varepsilon > 0$ *and a triple* $(P, Q, R) \in \mathcal{J}_0$ *such that at least one of the following two inequalities holds*: $((I - R)^{-1}P)(i, j) \geq \varepsilon$ *for all* $i,\ j \in [1, m]$, *or* $((I - R)^{-1}Q)(i, j) \geq \varepsilon$ *for all* $i,\ j \in [1, m]$.

Our proofs will be carried out under Condition **C2**. They can be modified to make them work also under Condition **C3**. Lemma 6 which is used in the proof of Theorem 1 is the main statement requiring a more careful treatment under Condition **C3** and the corresponding adjustments are not difficult. However, the proofs become more technical in this case, and we shall not do this in the present paper. If now vectors $p(x, \cdot)$ defining matrices (1.7) are $\mathbb{P}$-almost surely such that $p(x, 1) \geq \epsilon$ and $p(x, -1) \geq \epsilon$ for some $\epsilon > 0$, then it is easy to see that Condition **C3** is satisfied. We note also that if in addition the inequalities $p(x, m) \geq \epsilon$ and $p(x, -m) \geq \epsilon$ hold $\mathbb{P}$-almost surely, then also **C2** is satisfied.

For a triple of matrices $(P, Q, R) \in \mathcal{J}_0$ denote by $\pi = \pi_{(P,Q,R)} = (\pi_1, \ldots, \pi_m)$ a row vector with non-negative components such that

$$\pi(P + Q + R) = \pi \ \text{ and } \sum_{j=1}^{m} \pi_j = 1.$$

Note that the vector π is uniquely defined. Indeed, the equation for π can be rewritten as

$$\pi(I-R)\left((I-R)^{-1}P+(I-R)^{-1}Q\right)=\pi(I-R).$$

According to Condition **C2**, the stochastic matrix $(I-R)^{-1}P+(I-R)^{-1}Q$ has strictly positive elements (in fact they are $\geq 2\varepsilon$). Hence $\pi(I-R)$ is uniquely (up to a multiplication by a number) defined by the last equation and this implies the uniqueness of π.

Consider the following subset of $\mathcal{J}$:

$$\mathcal{J}_{al} \stackrel{\text{def}}{=} \{(P,Q,R)\in\mathcal{J} \,:\, \pi(P-Q)\mathbf{1}=0, \text{ where } \pi(P+Q+R)=\pi\}, \tag{1.8}$$

where obviously $\pi(P-Q)\mathbf{1}\equiv\sum_{i=1}^{m}\pi_i\sum_{j=1}^{m}(P(i,j)-Q(i,j))$. Note that $\mathcal{J}_{al}$ is an algebraic subsurface in $\mathcal{J}$.

We are now in a position to state the main result of this work:

Theorem 1. *Suppose that Condition* **C** *is satisfied, the random walk* $\xi(\cdot)=(X(\cdot),Y(\cdot))$ *is recurrent, and* $\mathcal{J}_0\not\subset\mathcal{J}_{al}$. *Then there is a sequence of random variables* $b_t(\omega)$, $t=1,2,\ldots$, *which converges weakly as* $t\to\infty$ *and such that for any* $\epsilon>0$,

$$\mathbb{P}\left\{\omega \,:\, Pr_\omega\left(|\frac{X(t)}{(\log t)^2}-b_t|\leq\epsilon\right)\geq 1-\epsilon\right\}\to 1 \text{ as } t\to\infty. \tag{1.9}$$

Remarks. The algebraic condition in this theorem requires a certain degree of non-degeneracy of the support $\mathcal{J}_0$ of the distribution of (P_n,Q_n,R_n). It may happen that relations (1.9) hold even when $\mathcal{J}_0\subset\mathcal{J}_{al}$. However Theorem 3 shows that there are important classes of environments where relations (1.9) (or (1.11)) hold if and only if this non-degeneracy condition is satisfied.

We now turn to the one-dimensional model. It should be mentioned right away that Theorem 2 is essentially a corollary of Theorem 1.

Denote by $\tilde{\mathcal{J}}$ the set of all $2m+1$-dimensional probability vectors:

$$\tilde{\mathcal{J}} \stackrel{\text{def}}{=} \{(p(j))_{-m\leq j\leq m} \,:\, p(\cdot)\geq 0 \text{ and } \sum_{j=-m}^{m}p(j)=1\}.$$

Remember that in this model the environment is a sequence of vectors: $\omega=(p(x,\cdot))_{-\infty<x<\infty}$, where $p(x,\cdot)\in\tilde{\mathcal{J}}$. Let $\tilde{\mathcal{J}}_0\subset\tilde{\mathcal{J}}$ be the support of the distribution of the random vector $p(0,\cdot)$. Finally, put

$$\tilde{\mathcal{J}}_{al} \stackrel{\text{def}}{=} \{p(\cdot)\in\tilde{\mathcal{J}} \,:\, \sum_{j=-m}^{m}jp(j)=0\}. \tag{1.10}$$

Theorem 2. *Suppose that:*
(a) $p(x,\cdot)$, $x\in\mathbb{Z}$, *is a sequence of i. i. d. vectors,*
(b) *there is an* $\varepsilon>0$ *such that* $p(0,1)\geq\varepsilon$, $p(0,-1)\geq\varepsilon$, $p(0,m)\geq\varepsilon$, *and* $p(0,-m)\geq\varepsilon$ *for any* $p(0,\cdot)\in\tilde{\mathcal{J}}_0$,
(c) *for* $\mathbb{P}$ *almost all environments* ω *the corresponding one-dimensional random walk* $\xi(\cdot)$ *is recurrent,*
(d) $\tilde{\mathcal{J}}_0\not\subset\tilde{\mathcal{J}}_{al}$.

Then there is a weakly converging sequence of random variables $b_t(\omega),\ t = 1, 2, \ldots$ *such that for any* $\epsilon > 0$,

$$\mathbb{P}\left\{\omega\ :\ Pr_\omega\left(|\frac{\xi(t)}{(\log t)^2} - b_t| \le \epsilon\right) \ge 1-\epsilon\right\} \to 1 \text{ as } t \to \infty. \tag{1.11}$$

Proof. Since the one-dimensional model reduces to a model on a strip, the result in question would follow if we could check that all conditions of Theorem 1 follow from those of Theorem 2.

It is obvious from formulae (1.7) that the i. i. d. requirement (Condition **C1**) follows from condition (a) of Theorem 2. We have already mentioned above that Condition **C2** follows from condition (b). The recurrence of the corresponding walk on a strip is also obvious.

Finally, condition (d) implies the algebraic condition of Theorem 1. Indeed, formulae (1.7) show that matrices P_n, Q_n, R_n are defined by probability vectors $p(nm+i,\cdot) \in \tilde{\mathcal{J}}_0$, where $1 \le i \le m$. Put $n = 0$ and choose all these vectors to be equal to each other, say $p(i,\cdot) = p(\cdot) \in \tilde{\mathcal{J}}_0$, where $1 \le i \le m$. A direct check shows that the triple of matrices (P, Q, R) built from this vector has the property that $P + Q + R$ is double-stochastic and irreducible (irreducibility follows from the conditions $p(1) \ge \varepsilon$ and $p(-1) \ge \varepsilon$). Hence the only probability vector π satisfying $\pi(P+Q+R) = \pi$ is given by $\pi = (m^{-1}, \ldots, m^{-1})$. One more direct calculation shows that in this case

$$m\pi(P - Q)\mathbf{1} = \sum_{j=-m}^{m} jp(j).$$

Hence the condition $\mathcal{J}_0 \not\subset \mathcal{J}_{al}$ of Theorem 1 is satisfied if there is at least one vector $p(\cdot) \in \tilde{\mathcal{J}}_0$ such that $\sum_{j=-m}^{m} jp(j) \ne 0$. □

We conclude this section with a theorem which shows, among other things, that the algebraic condition of Theorem 2 is also necessary for having (1.11). This theorem does not require independence as such but in a natural sense it finalizes the classification of the one-dimensional recurrent RWs with bounded jumps in the i. i. d. environments.

Theorem 3. *Consider a one-dimensional RW and suppose that*
(a) $p(x,\cdot),\ x \in \mathbb{Z}$, *is a strictly stationary ergodic sequence of vectors,*
(b) *there is an* $\varepsilon > 0$ *such that* $p(0,1) \ge \varepsilon$ *and* $p(0,-1) \ge \varepsilon$ *for any* $p(0,\cdot) \in \tilde{\mathcal{J}}_0$,
(c) $\tilde{\mathcal{J}}_0 \subset \tilde{\mathcal{J}}_{al}$, *that is*

$$\sum_{j=-m}^{m} jp(j) = 0 \text{ for any } p(\cdot) \in \tilde{\mathcal{J}}_0 .$$

Then:
(i) *The random walk* $\xi(\cdot)$ *is asymptotically normal in every(!) environment* $\omega = (p(x,\cdot))_{-\infty<x<\infty}$.
(ii) *There is a* $\sigma > 0$ *such that for* $\mathbb{P}$*-a. e.* ω,

$$\lim_{t\to\infty} Pr_\omega\left\{\frac{\xi(t)}{\sqrt{t}} \le x\right\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{u^2}{2\sigma^2}}du, \tag{1.12}$$

where x *is any real number and the convergence in* (1.12) *is uniform in* x.

Remarks about the proof of Theorem 3. The condition of this theorem implies that $\xi(t)$ is a martingale:

$$E_\omega(\xi(t) - \xi(t-1) \,|\, \xi(t-1) = k) = \sum_{j=-m}^{m} jp(k, j) = 0,$$

where E_ω denotes the expectation with respect to the probability measure Pr_ω on the space of trajectories of the random walk (we assume that $\xi(0) = 0$). Let $U_n = \xi(n) - \xi(n-1)$ and put

$$\sigma_n^2 \stackrel{\text{def}}{=} E_\omega(U_n^2 \,|\, \xi(n-1)) = \sum_{j=-m}^{m} j^2 p(\xi(n-1), j).$$

Obviously $\varepsilon \le \sigma_n^2 \le m^2$, where ε is the same as in Theorem 3. Next put $V_n^2 \stackrel{\text{def}}{=} \sum_{j=1}^{n} \sigma_j^2$ and $s_n^2 \stackrel{\text{def}}{=} E_\omega(V_n^2) = E_\omega(\xi(n)^2)$. It is useful to note that $n\varepsilon \le V_n^2,\ s_n^2 \le nm^2$. Let $T_t = \inf\{n : V_n^2 \ge t\}$.

Statement (i) of Theorem 3 is a particular case of a much more general theorem of Drogin who in particular proves that $t^{-1/2}\xi(T_t)$ converges weakly to a standard normal random variable. We refer to [12], p. 98 for more detailed explanations.

Statement (ii) of Theorem 3 is similar to a well known result by Lawler [16]. The main ingredient needed for proving (ii) is the following claim:

$$\text{The limit } \lim_{n\to\infty} n^{-1} V_n^2 = \lim_{n\to\infty} n^{-1} s_n^2 \text{ exists for } \mathbb{P}\text{-almost all } \omega. \tag{1.13}$$

Once this property of the variance of $\xi(\cdot)$ is established, (ii) becomes a corollary of Brown's Theorem (see Theorems 9 and 10 in Appendix or Theorem 4.1 in [12]).

However proving (1.13) is not an entirely straightforward matter. The proof we are aware of uses the approach known under the name "environment viewed from the particle". This approach was used in [16] for proving properties of variances similar to (1.13); unfortunately, the conditions used in [16], formally speaking, are not satisfied in our case. Fortunately, Zeitouni in [20] found the way in which Lawler's result can be extended to more general martingale-type random walks in random environments which include our case. □

2. Preparatory Results

2.1. Elementary corollaries of Condition **C**. We start with several elementary observations following from **C2**. Lemma 3 and a stronger version of Lemma 1 can be found in [1]. Lemmas 2 and 4 are borrowed from [6].

Lemma 1. *If Condition* **C2** *is satisfied then for $\mathbb{P}$-almost every environment ω the whole phase space $\mathbb{S}$ of the Markov chain $\xi(t)$ constitutes the (only) communication class of this chain.*

Proof. Fix an environment ω and consider matrices

$$\tilde{P}_n \stackrel{\text{def}}{=} (I - R_n)^{-1} P_n,\ \tilde{Q}_n \stackrel{\text{def}}{=} (I - R_n)^{-1} Q_n.$$

Remark that $\tilde{P}_n(i, j)$ is the probability that the random walk ξ starting at (n, i) would reach $(n + 1, j)$ at the time of its first exit from layer n; the probabilistic meaning of $\tilde{Q}_n(i, j)$ is defined similarly. $\tilde{P}_n(i, j) \geq \varepsilon > 0$ and $\tilde{Q}_n(i, j) \geq \varepsilon > 0$ because of Condition **C2**. It is now obvious that a random walk $\xi(\cdot)$ starting from any $z \in \mathbb{S}$ would reach any $z_1 \in \mathbb{S}$ with a positive probability. □

Matrices of the form $(I - R - Q\psi)^{-1}$, $(I - R - Q\psi)^{-1}P$, and $(I - R - Q\psi)^{-1}Q$ arise in the proofs of many statements below. We shall list several elementary properties of these matrices.

Lemma 2. *If Condition* **C2** *is satisfied,* $(P, Q, R) \in \mathcal{J}_0$ *and* ψ *is any stochastic matrix, then there is a constant* C *depending only on* ε *and* m *such that*

$$\left\| (I - R - Q\psi)^{-1} \right\| \leq C. \tag{2.1}$$

Proof. Note first that $||R^l|| \leq 1 - \varepsilon$ implies that for some C_1 uniformly in R,

$$\left\| (I - R)^{-1} \right\| \leq \sum_{k=0}^{\infty} \left\| R^k \right\| \leq C_1.$$

Next, it follows from $(P + Q + R)\mathbf{1} = \mathbf{1}$ that $(I - R)^{-1}P\mathbf{1} + (I - R)^{-1}Q\mathbf{1} = \mathbf{1}$ and $(I - R)^{-1}Q\mathbf{1} = \mathbf{1} - (I - R)^{-1}P\mathbf{1}$. Condition **C2** implies that $(I - R)^{-1}P\mathbf{1} \geq m\varepsilon\mathbf{1}$. Hence

$$\left\| (I - R)^{-1}Q \right\| = \left\| (I - R)^{-1}Q\mathbf{1} \right\| = \left\| \mathbf{1} - (I - R)^{-1}P\mathbf{1} \right\| \leq 1 - m\varepsilon.$$

Similarly, $\left\| (I - R)^{-1}P \right\| \leq 1 - m\varepsilon$. Hence

$$\begin{aligned} \left\| (I - R - Q\psi)^{-1} \right\| &= \left\| (I - (I - R)^{-1}Q\psi)^{-1}(I - R)^{-1} \right\| \\ &\leq \left(1 - \left\| (I - R)^{-1}Q\psi \right\| r\right)^{-1} \left\| (I - R)^{-1} \right\| \leq C_1 m^{-1}\varepsilon^{-1} \equiv C. \end{aligned}$$

Lemma is proved. □

Lemma 3. *([1]). If Condition* **C2** *is satisfied,* $(P, Q, R) \in \mathcal{J}$, *and* ψ *is a stochastic matrix, then* $(I - R - Q\psi)^{-1}P$ *is also stochastic.*

Proof. We have to check that $(I - R - Q\psi)^{-1}P\mathbf{1} = \mathbf{1}$ which is equivalent to $P\mathbf{1} = (I - Q\psi - R)\mathbf{1} \Leftrightarrow (P + Q\psi + R)\mathbf{1} = \mathbf{1}$. Since $\psi\mathbf{1} = \mathbf{1}$ and $P + Q + R$ is stochastic, the result follows. □

Lemma 4. *Suppose that Condition* **C2** *is satisfied and* $(P, Q, R) \in \mathcal{J}_0$ *and let a matrix* $\varphi \geq 0$ *be such that* $\varphi\mathbf{1} \leq \mathbf{1}$. *Then*

$$((I - R - Q\varphi)^{-1}P)(i, j) \geq \varepsilon \;\; \text{and} \;\; ((I - R - Q\varphi)^{-1}Q)(i, j) \geq \varepsilon. \tag{2.2}$$

Proof. $(I - R - Q\varphi)^{-1}P \geq (I - R)^{-1}P$ and $(I - R - Q\varphi)^{-1}Q \geq (I - R)^{-1}Q$. □

2.2. Random transformations, related Markov chains, Lyapunov exponents, and recurrence criteria. The purpose of this section is to introduce objects listed in its title. These objects shall play a major role in the proofs of our main results. They shall also allow us to state the main results from [1] in the form which is suitable for our purposes.

Random transformations and related Markov chains. Let Ψ be the set of stochastic $m \times m$ matrices, $\mathbb{X}$ be the set of unit vectors with non-negative components, and $\mathrm{M} \overset{\text{def}}{=} \Psi \times \mathbb{X}$ the direct product of these two sets. Define a distance $\rho(\cdot, \cdot)$ on M by

$$\rho((\psi, x), (\psi', x')) \overset{\text{def}}{=} ||\psi - \psi'|| + ||x - x'||. \tag{2.3}$$

For any triple $(P, Q, R) \in \mathcal{J}_0$ denote by $g \equiv g_{(P,Q,R)}$ a transformation

$$g : \mathrm{M} \mapsto \mathrm{M}, \text{ where } g.(\psi, x) \overset{\text{def}}{=} ((I - R - Q\psi)^{-1}P\,,\ ||Bx||^{-1}Bx), \tag{2.4}$$

and

$$B \equiv B_{(P,Q,R)}(\psi) \overset{\text{def}}{=} (I - R - Q\psi)^{-1}Q. \tag{2.5}$$

The fact that g maps M into itself follows from Lemma 3.

Remarks. Here and in the sequel the notation $g.(\psi, x)$ is used instead of $g((\psi, x))$ and the dot is meant to replace the brackets and to emphasize the fact that g maps (ψ, x) into another pair from M. In fact this notation is often used in the theory of products of random matrices, e. g. $B.x \overset{\text{def}}{=} ||Bx||^{-1}Bx$; we thus have extended this tradition to another component of g.

If $\omega \in \Omega$ is an environment, $\omega = (\omega_n)_{-\infty<n<\infty}$, where $\omega_n \overset{\text{def}}{=} (P_n, Q_n, R_n) \in \mathcal{J}_0$, then (2.4) allows us to define a sequence $g_n \equiv g_{\omega_n}$ of random transformations of M. Given the sequence g_n, we define a Markov chain with a state space $\mathcal{J}_0 \times \mathrm{M}$. To this end consider an $a \in \mathbb{Z}$, and a $(\psi_a, x_a) \in \mathrm{M}$ and put for $n \geq a$,

$$(\psi_{n+1}, x_{n+1}) \overset{\text{def}}{=} g_n.(\psi_n, x_n) \equiv ((I - R_n - Q_n\psi_n)^{-1}P_n\,,\ \|B_n x_n\|^{-1}B_n x_n), \tag{2.6}$$

where we use a concise notation for matrices defined by (2.5):

$$B_n \overset{\text{def}}{=} B_{\omega_n}(\psi_n) \equiv B_{(P_n,Q_n,R_n)}(\psi_n). \tag{2.7}$$

Theorem 4. *Suppose that Condition* **C** *is satisfied. Then:*
a) *For* $\mathbb{P}$*-a.e. sequence* ω *the following limits exist:*

$$\zeta_n \overset{\text{def}}{=} \lim_{a\to-\infty} \psi_n, \quad y_n \overset{\text{def}}{=} \lim_{a\to-\infty} x_n, \tag{2.8}$$

and (ζ_n, y_n) *does not depend on the choice of the sequence* (ψ_a, y_a)*. Furthermore, the convergence in (2.8) is uniform in* (ψ_a, x_a)*.*
b) *The sequence of pairs* $(\zeta_n, y_n) \equiv (\zeta_n(\omega), y_n(\omega))$ $-\infty < n < \infty$*, is the unique sequence of elements from* M *which satisfy the following infinite system of equations*

$$(\zeta_{n+1}, y_{n+1}) = \Big((I - R_n - Q_n\zeta_n)^{-1}P_n\,,\ ||A_n(\omega)y_n||^{-1}A_n(\omega)y_n\Big), \quad n \in \mathbb{Z}, \tag{2.9}$$

where

$$A_n \equiv A_n(\omega) \overset{\text{def}}{=} (I - R_n - Q_n\zeta_n)^{-1}Q_n. \tag{2.10}$$

c) *The enlarged sequence* (ω_n, ζ_n, y_n)*,* $-\infty < n < \infty$*, forms a stationary and ergodic Markov chain with components* ω_n *and* (ζ_n, y_n) *being independent of each other.*

Proof. The first relation in (2.8) is the most important statement of our theorem and it also is the main content of Theorem 1 in [1]; it thus is known.

The main difference between this theorem and Theorem 1 from [1] is that here we consider the extended sequence $(\psi_n, x_n),\ n \geq a$, rather than just $(\psi_n),\ n \geq a$. The proof of the second relation in (2.8) is based on two observations. First note that the first relation in (2.8) implies that $\lim_{a\to-\infty} B_n = A_n$. Next, it follows from the definition of the sequence x_n that

$$x_n = \|B_{n-1}\dots B_a x_a\|^{-1}\, B_{n-1}\dots B_a x_a. \tag{2.11}$$

Estimates (2.1) and (2.2) imply that $\min_{i_1,i_2,i_3,i_4} B_k^{-1}(i_1,i_2)B_k(i_3,i_4) \geq \bar{\varepsilon}$ for some $\bar{\varepsilon} > 0$ and hence also $\min_{i_1,i_2,i_3,i_4} A_k^{-1}(i_1,i_2)A_k(i_3,i_4) \geq \bar{\varepsilon}$. It is well known (and can be easily derived from Lemma 15) that these inequalities imply the existence of

$$\lim_{a\to-\infty} \|A_nA_{n-1}\dots A_a x_a\|^{-1}\, A_nA_{n-1}\dots A_a x_a$$

and this limit does not depend on the choice of the sequence $x_a \geq 0$, $||x_a|| = 1$. Combining these two limiting procedures we obtain the proof of the second relation in (2.8).

Part b) of the theorem is proved exactly as part b) of Theorem 1 from [1].

The Markov chain property and the independence claimed in part c) are obvious corollaries of the independence of the triples (P_n, Q_n, R_n). And, finally, the ergodicity of the sequence (ω_n, ζ_n, y_n) is due to the fact that the sequence ω_n is ergodic and the (ζ_n, y_n) is a function of $(\omega_k)_{k\leq n-1}$. □

Remarks. The proof of Theorem 1 in [1] was obtained under much less restrictive assumptions than those listed in Condition **C** of this work. In particular, the i. i. d. condition which we impose on our environments (rather than having them just stationary and ergodic) is unimportant for parts a) and b) of Theorem 4 as well as for Theorem 5. However, the i. i. d. property is important for the proof of our main results.

The top Lyapunov exponent of products of matrices A_n and the recurrence criteria. The top Lyapunov exponent of products of matrices A_n will be denoted by λ and it is defined by

$$\lambda \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \log \|A_nA_{n-1}\dots A_1\|\,. \tag{2.12}$$

The existence of the limit in (2.12) with $\mathbb{P}$-probability 1 and the fact that λ does not depend on ω is an immediate corollary of Kingman's sub-additive ergodic theorem; it was first proved in [5]. The Furstenberg formula states that

$$\lambda = \int_{\mathcal{J}_0\times\mathrm{M}} \log \left\|(I - R - Q\zeta)^{-1}Qy\right\| \mu(dg)\nu(d(\zeta, y)), \tag{2.13}$$

where $\nu(d(\zeta, y))$ is the invariant measure of the Markov chain (2.6) and $\mu(dg)$ is the distribution of the set of triples (P, Q, R) supported by $\mathcal{J}_0$ (defined in Sect. 1.3). We use the shorter notation dg rather than $d(P, Q, R)$ because, as we have seen above, every triple $(P, Q, R) \in \mathcal{J}_0$ defines a transformation g. Besides, this notation is consistent with the one used in Sect. 4.1.3.

We remark that a proof of (2.12) and (2.13) will be given in Sect. 2.4 as a natural part of the proof of the invariance principle for the sequence of random variables $\log\|A_nA_{n-1}\dots A_1\|$.

We finish this section by quoting the recurrence criteria proved in [1].

Theorem 5. *Suppose that Condition* **C** *is satisfied. Then*
a) $\lambda \gtrless 0$ *if and only if for* $\mathbb{P}$*-a.e. environment* ω *one has* (*respectively*)

$$\lim_{t\to\infty} \xi(t) = \mp\infty \ \ Pr_\omega\text{-almost surely.}$$

b) $\lambda = 0$ *if and only if for* $\mathbb{P}$*-a.e.* ω *the RW* $\xi(\cdot)$ *is recurrent, that is*

$$\limsup_{t\to\infty} \xi(t) = +\infty \ \ \text{and} \ \ \liminf_{t\to\infty} \xi(t) = -\infty \ \ Pr_\omega\text{-almost surely.}$$

2.3. One algebraic corollary of Theorems 4 and 5. Theorems 4 and 5 combined with a simple probabilistic observation lead to an algebraic result which plays a very important role in the proof of our algebraic condition.

Suppose that the matrices (P_n, Q_n, R_n) do not depend on n: $(P_n, Q_n, R_n) \equiv (P, Q, R)$, and the triple (P, Q, R) satisfies Condition **C2**. In this case relations (2.8) mean that $\zeta_n = \zeta$ and $y_n = y$, where ζ is a unique stochastic matrix and $y \geq 0$ a unique unit vector such that

$$\zeta = (I - R - Q\zeta)^{-1}P, \ \ \text{and} \ \ Ay = e^{\lambda}y, \tag{2.14}$$

where the matrix A is defined by

$$A \stackrel{\text{def}}{=} (I - R - Q\zeta)^{-1}Q.$$

Theorem 5 now states that a random walk in a constant environment is recurrent if $\lambda = 0$, transient to the right if $\lambda < 0$, and transient to the left if $\lambda > 0$.

But the fact that the random environment does not depend on n allows one to analyse the recurrence and transience properties of the random walk in a way which is much more straightforward than the one offered by Theorems 4 and 5.

Namely, suppose that $\xi(t) = (X(t), Y(t)) = (k, i)$. Then the conditional probability $Pr\{Y(t) = j \,|\, \xi(t-1) = (k, i)\} = P(i, j) + Q(i, j) + R(i, j)$ does not depend on $X(t-1)$ and thus the second coordinate of this walk is a Markov chain with a state space $(1, \ldots, m)$ and a transition matrix $P + Q + R$. Hence, if $\pi = (\pi_1, \ldots \pi_m)$ is a probability vector such that $\pi(P + Q + R) = \pi$ then π_i is the frequency of visits by the RW to the sites $(\cdot, i)$ of the strip.

Consider next the displacement $\eta(t) \stackrel{\text{def}}{=} X(t) - X(t-1)$ of the coordinate X of the walk which occurs between times $t-1$ and t. The random variable $\eta(t)$ takes values $1, -1$, or 0 and the following conditional distribution of the pair $(\eta(t), Y(t))$ is given by $Pr\{(\eta(t), Y(t)) = (1, j) \,|\, \xi(t-1) = (k, i)\} = P(i, j)$, $Pr\{(\eta(t), Y(t)) = (-1, j) \,|\, \xi(t-1) = (k, i)\} = Q(i, j)$, and $Pr\{(\eta(t), Y(t)) = (0, j) \,|\, \xi(t-1) = (k, i)\} = R(i, j)$. It is essential that this distribution depends only on i (and not on k) and thus this pair forms a time-stationary Markov chain. Let us denote by $E_{(k,i)}$ the corresponding conditional expectation with conditioning on $(\eta(t-1), Y(t-1)) = (k, i)$, $-1 \leq k \leq 1, \ 1 \leq m$. We then have

$$E_{(k,i)}(\eta(t)) = \sum_{j=1}^{m} P(i, j) - \sum_{j=1}^{m} Q(i, j),$$

and the expectation of the same random variable with respect to the stationary distribution is thus given by $\sum_{i=1}^{m} \pi_i \sum_{j=1}^{m}(P(i, j) - Q(i, j))$. Applying the law of large numbers for Markov chains to the sequence $\eta(t)$ we obtain that with Pr-probability 1,

$$\lim_{t\to\infty} t^{-1}X(t) = \lim_{t\to\infty} t^{-1}\sum_{k=1}^{t}\eta(k) = \sum_{i=1}^{m}\pi_i\sum_{j=1}^{m}(P(i,j)-Q(i,j)),$$

and this limit is independent of the $\xi(0)$. Since this result is equivalent to the statements of Theorems 4 and 5, we obtain the following

Lemma 5. *Suppose that* (P, Q, R) *satisfies Condition* **C2**. *Then* $(\zeta, x) \in \mathrm{M}$ *satisfies Eq.* (2.14) *with* $\lambda = 0$ *if and only if*

$$\sum_{i=1}^{m}\pi_i\sum_{j=1}^{m}(P(i,j)-Q(i,j)) = 0. \tag{2.15}$$

Moreover $\lambda > 0$ *if and only if* $\sum_{i=1}^{m}\pi_i\sum_{j=1}^{m}(P(i,j)-Q(i,j)) < 0$ *(and thus* $\lambda < 0$ *if and only if* $\sum_{i=1}^{m}\pi_i\sum_{j=1}^{m}(P(i,j)-Q(i,j)) > 0$*).*

2.4. The CLT and the invariance principle for S_n's. The main goal of this section is to prove an invariance principle (IP) (and a CLT) for the sequence

$$S_n \stackrel{\text{def}}{=} \log\|B_n\dots B_1 x_1\| - n\lambda, \tag{2.16}$$

where matrices B_n are defined by (2.7) and λ is given by (2.13). Obviously, S_n depends on $(\psi_1, x_1) \in \mathrm{M}$. We shall prove that in fact the IP (and the CLT) are satisfied uniformly in $(\psi_1, x_1) \in \mathrm{M}$. Moreover, exactly one of the two things takes place if the random walk is recurrent: either the asymptotic behaviour of S_n is described by a non-degenerate Wiener process, or the support of the distribution of matrices (P, Q, R) belongs to an algebraic manifold defined by Eq. (1.8).

To make these statements precise we first recall one of the definitions of the invariance principle associated with a general random sequence $S_n = \sum_{k=1}^{n} f_k$, with the convention $S_0 = 0$. Let $\{C[0,1], \mathcal{B}, P_W\}$ be the probability space where $C[0,1]$ is the space of continuous functions with the sup norm topology, $\mathcal{B}$ being the Borel σ-algebra generated by open sets in $C[0,1]$, and P_W the Wiener measure. Define for $t \in [0,1]$ a sequence of random functions $v_n(t)$ associated with the sequence S_n. Namely, put

$$v_n(t) \stackrel{\text{def}}{=} n^{-\frac{1}{2}}\left(S_k + f_{k+1}(tn-k)\right) \quad \text{if } k \le tn \le k+1,\ k = 0, 1, \dots, n-1. \tag{2.17}$$

For a $\sigma > 0$ let $\{\mathbb{P}_n^\sigma\}$ be the sequence of probability measures on $\{C[0,1], \mathcal{B}\}$ determined by the distribution of $\{\sigma^{-1}v_n(t),\ 0 \le t \le 1\}$.

Definition. *A random sequence S_n satisfies the invariance principle with parameter $\sigma > 0$ if $\mathbb{P}_n^\sigma \to P_W$ weakly as $n \to \infty$. If the sequence S_n depends on (another) parameter, e.g. z_1, then we say that S_n satisfies the invariance principle with parameter $\sigma > 0$ uniformly in z_1 if for any continuous functional on $\mathfrak{f} : C[0,1] \mapsto \mathbb{R}$ one has: $\mathbb{E}_n^\sigma(\mathfrak{f}) \to E_W(\mathfrak{f})$ uniformly in z_1 as $n \to \infty$. Here $\mathbb{E}_n$ and E_W are expectations with respect to the relevant probabilities.*

Let us state the invariance principle for the sequence S_n given by (2.16). Note that in this case

$$S_n = \sum_{k=1}^{n} (\log \|B_k x_k\| - \lambda), \text{ where } x_k = \|B_{k-1}x_{k-1}\|^{-1} B_{k-1}x_{k-1}, \ k \geq 2. \tag{2.18}$$

Put $z_n = (\psi_n, x_n)$ and $f_n = f(g_n, z_n)$, where the function f is defined on the set of pairs $(g, z) \equiv ((P, Q, R), (\psi, x))$ by

$$f(g, z) \stackrel{\text{def}}{=} \log \left\| (I - R - Q\psi)^{-1} Qx \right\| - \lambda. \tag{2.19}$$

Obviously in these notations $S_n = \sum_{k=1}^{n} f_k$. Denote by $\mathfrak{A}$ the Markov operator associated with the Markov chain $z_{n+1} = g_n.z_n$ defined by (2.6): if F is a function defined on the state space $\mathcal{J}_0 \times \mathrm{M}$ of this chain then

$$(\mathfrak{A}F)(g, z) \stackrel{\text{def}}{=} \int_{\mathcal{J}_0 \times \mathrm{M}} F(g', g.z)\mu(dg').$$

Using these notations we write $\nu(dz)$ (rather than $\nu(d(\psi, x))$) for the invariant measure of the chain z_n and we denote by $\mathrm{M}_0 \subset \mathrm{M}$ the support of $\nu(dz)$.

Theorem 6. *Suppose that Condition* **C** *is satisfied and the function f is defined by* (2.19). *Then:*
(i) *The equation*

$$F(g, z) - (\mathfrak{A}F)(g, z) = f(g, z) \tag{2.20}$$

has a unique solution $F(g, z)$ which is continuous on $\mathcal{J}_0 \times \mathrm{M}_0$ and

$$\int_{\mathcal{J}_0 \times \mathrm{M}} F(g, z)\mu(dg)\nu(dz) = 0.$$

Denote by

$$\sigma^2 = \int_{\mathcal{J}_0 \times \mathrm{M}_0} (\mathfrak{A}F^2 - (\mathfrak{A}F)^2)(g, y)\mu(dg)\nu(dy).$$

(ii) *If $\sigma > 0$ then $\frac{S_n}{\sigma\sqrt{n}}$ converges in law towards the standard Gaussian distribution $N(0, 1)$ and the sequence S_n satisfies the invariance principle with parameter σ uniformly in $(\psi_1, x_1) \in \mathrm{M}$.*
(iii) *If $\sigma = 0$, then the function $F(g, y)$ depends only on y and for every $(g, y) \in \mathcal{J}_0 \times \mathrm{M}_0$ one has*

$$f(g, y) = F(y) - F(g.y). \tag{2.21}$$

(iv) *If $\sigma = 0$ and $\lambda = 0$ then*

$$\mathcal{J}_0 \subset \mathcal{J}_{al}, \tag{2.22}$$

with $\mathcal{J}_{al}$ given by (1.8).

Proof. Statements (i), (ii), and (iii) of our theorem follow from Theorem 12. In order to be able to apply Theorem 12 we have to show that the sequence of random transformations g_n has the so called contraction property. Lemma 6 establishes this property. Relation (2.22) is then derived from (2.21) and one more general property of Markov chains generated by products of contracting transformations (Lemma 8).

Lemma 6. *Suppose that Condition* **C** *is satisfied and let*

$$(\psi_{n+1}, x_{n+1}) = g_n.(\psi_n, x_n), \quad (\psi'_{n+1}, x'_{n+1}) = g_n.(\psi'_n, x'_n), \quad n \geq 1,$$

be two sequences from M. *Then there is a* $c,\ 0 \leq c < 1$, *such that for any* (ψ_1, x_1), $(\psi'_1, x'_1) \in$ M,

$$\rho\left((\psi_n, x_n), (\psi'_n, x'_n)\right) \leq \text{const}\, c^n, \tag{2.23}$$

where $\rho(\cdot, \cdot)$ *is defined by* (2.3).

Proof of Lemma 6. We shall first prove that there is a $c_0 < 1$ such that $||\psi_n - \psi'_n|| \leq$ const c_0^n. The control of the x-component would then follow from this result.

Let us introduce a sequence of $m \times m$ matrices $\varphi_n, n \geq 1$, which we define recursively: $\varphi_1 = 0$ and

$$\varphi_{n+1} = (I - R_n - Q_n\varphi_n)^{-1} P_n, \quad \text{if } n \geq 1. \tag{2.24}$$

Remarks. Matrices φ_n and ψ_n were defined in a purely analytic way. Their probabilistic meaning is well known (see [1]) and shall also be discussed in Sect. 3.

Put $\Delta_k \stackrel{\text{def}}{=} \psi_k - \varphi_k$. To control the ψ-part of the sequence (ψ_n, x_n) we need the following

Lemma 7. *Suppose that Condition* **C** *is satisfied. Then there is a* $c_0,\ 0 \leq c_0 < 1$, *such that for any stochastic matrix* $\psi_1 \in \Psi$ *the matrix elements of the corresponding* Δ_{n+1} *are of the following form:*

$$\Delta_{n+1}(i, j) = \alpha_n(i)c_n(j) + \tilde{\epsilon}_n(i, j). \tag{2.25}$$

Here $\alpha_n(i)$ *and* $c_n(j)$ *depend only on the sequence* $(P_j, Q_j, R_j), 1 \leq j \leq n$; *the matrix* $\tilde{\epsilon}_n = (\tilde{\epsilon}_n(i, j))$ *is a function of* ψ_1 *and of the sequence* (P_j, Q_j, R_j), $1 \leq j \leq n$, *satisfying* $||\tilde{\epsilon}_n|| \leq C_1 c_0^n$ *for some constant* C_1.

Corollary. If Condition **C** holds then

$$||\psi_{n+1} - \psi'_{n+1}|| \leq 2C_1\, c_0^n. \tag{2.26}$$

Proof of Corollary. Consider a sequence ψ'_n which differs from ψ_n in that the starting value for recursion (2.6) is ψ'_1. Put $\Delta'_k \stackrel{\text{def}}{=} \psi'_k - \varphi_k$. Applying the result of Lemma 7 to Δ'_{n+1} we obtain:

$$\Delta'_{n+1}(i, j) = \alpha_n(i)c_n(j) + \tilde{\epsilon}'_n(i, j). \tag{2.27}$$

It follows from (2.25), (2.27), and the definition of Δ_{n+1} and Δ'_{n+1} that $||\psi_{n+1} - \psi'_{n+1}|| = ||\Delta_{n+1} - \Delta'_{n+1}|| \leq ||\tilde{\epsilon}_n|| + ||\tilde{\epsilon}'_n|| \leq 2C_1\, c_0^n$. □

Proof of Lemma 7. The main idea of this proof is the same as that of the proof of Theorem 1 from [1]. A very minor difference is that here we have to control the behaviour of ψ_n when n is growing while ψ_1 is fixed; in [1] n was fixed while the starting point of the chain was tending to $-\infty$. A more important difference is that here we state the exponential speed of convergence of certain sequences and present the corresponding quantities in a relatively explicit way while in [1] the speed of convergence was not very essential (even though the exponential character of convergence had been clear already then).

To start, note that it follows from (2.6) and (2.24) that

$$\begin{aligned}\Delta_{n+1} &= ((I - R_n - Q_n\psi_n)^{-1} - (I - R_n - Q_n\varphi_n)^{-1})P_n \\ &= (I - R_n - Q_n\psi_n)^{-1}Q_n\Delta_n(I - R_n - Q_n\varphi_n)^{-1}P_n = B_n\Delta_n\varphi_{n+1}.\end{aligned} \tag{2.28}$$

Iterating (2.28), we obtain

$$\Delta_{n+1} = B_n \dots B_1\Delta_1\varphi_2 \dots \varphi_{n+1} \equiv B_n \dots B_1\psi_1\varphi_2 \dots \varphi_{n+1}. \tag{2.29}$$

It follows from Lemma 4 that $\varphi_n\mathbf{1} \le \mathbf{1}$. The matrix elements of the matrices φ_n, $n \ge 2$, are strictly positive and, moreover, according to estimates (2.2) we have: $\varphi_n(i,j) \ge \varepsilon$ (and hence also $\varphi_n(i,j) \le 1 - (m-1)\varepsilon$). We are in a position to apply to the product of matrices φ_n the presentation derived in Lemma 15 (with a_n's replaced by φ_n's). By the first formula in (4.16), we have:

$$\varphi_2 \dots \varphi_{n+1} = D_n[(c_n(1)\mathbf{1}, \dots, c_n(m)\mathbf{1}) + \phi_n],$$

where D_n is a diagonal matrix, $c_n(j) \ge \delta$ with $\sum_{j=1}^m c_n(j) = 1$, and $\|\phi_n\| \le (1 - m\delta)^{n-1}$ with $\delta > 0$ (and of course $m\delta < 1$). One can easily see that $\delta \ge m^{-1}\varepsilon^2$ (this follows from (4.15) and the above estimates for $\varphi_n(i,j)$). We note also that the estimate for $c_n(j)$ follows from (4.17) and (4.18).

Put $c_0 = 1 - m\delta$ and let $\mathcal{B}_n \stackrel{\text{def}}{=} B_n \dots B_1\Delta_1 D_n$. We then have

$$\Delta_{n+1} = \mathcal{B}_n[(c_n(1)\mathbf{1}, \dots, c_n(m)\mathbf{1}) + \phi_n], \tag{2.30}$$

and thus $\Delta_{n+1}(i,j) = c_n(j)\sum_{k=1}^m \mathcal{B}_n(i,k)\left(1 + \frac{\phi_n(k,j)}{c_n(j)}\right)$. But all $\mathcal{B}_n(i,k) > 0$ and $\max_{k,j}|\phi_n(k,j)|c_n^{-1}(j) \le \text{const}\, c_0^n$. Hence

$$\frac{\Delta_{n+1}(i,l)}{\Delta_{n+1}(i,j)} = \frac{c_n(l)}{c_n(j)} + \epsilon_n(i,j,l), \tag{2.31}$$

where $|\epsilon_n(i,j,l)| < Cc_0^n$ with C being some constant. It follows from (2.31) that

$$(\Delta_{n+1}(i,j))^{-1}\sum_{l=1}^m \Delta_{n+1}(i,l) = \frac{1}{c_n(j)} + \epsilon_n(i,j).$$

On the other hand remember that

$$\sum_{l=1}^m \Delta_{n+1}(i,l) = \sum_{l=1}^m \psi_{n+1}(i,l) - \sum_{l=1}^m \varphi_{n+1}(i,l) = 1 - \sum_{l=1}^m \varphi_{n+1}(i,l) \stackrel{\text{def}}{=} \alpha_n(i).$$

Comparing these two expressions we obtain that

$$\Delta_{n+1}(i,j) = \alpha_n(i)c_n(j) + \tilde{\epsilon}_n(i,j), \tag{2.32}$$

where $|\tilde{\epsilon}_n(i,j)| \le C_1c_0^n$. Lemma 7 is proved. □

We now turn to the difference $||x_{n+1} - x'_{n+1}||$. Let us denote by b_n the transformation of the set $\mathbb{X}$ of unit non-negative vectors defined by

$$b_n(x) = ||B_n x||^{-1} B_n x, \quad \text{where } B_n = (I - R_n - Q_n\psi_n)^{-1} Q_n, \tag{2.33}$$

and ψ_n are the same as above. The sequence b'_n is defined in a similar way with the only difference that ψ_n is replaced by ψ'_n. Inequality (2.26) implies that for some C_2,

$$\bar{\rho}(b_n, b'_n) \stackrel{\text{def}}{=} \sup_{x \in \mathbb{X}} ||b_n(x) - b'_n(x)|| \le C_2 c_0^n.$$

A very general and simple Lemma 16 from the Appendix now implies that

$$||x_{n+1} - x'_{n+1}|| \le C(\epsilon)(c_0 + \epsilon)^n (1 + ||x_1 - x'_1||)$$

and this proves Lemma 6. □

We can now easily prove the existence of the limit in (2.12) as well as Furstenberg's formula (2.13) for λ. To this end note that

$$\bar{S}_n(\zeta_1, \mathbf{1}) \stackrel{\text{def}}{=} \log ||A_n \dots A_1|| = \log ||A_n \dots A_1 \mathbf{1}|| = \sum_{k=1}^{n} f(g_k, z_k), \tag{2.34}$$

where the notation is chosen so that to emphasize the dependence of the sum $\bar{S}_n(\zeta_1, \mathbf{1})$ on initial values $x_1 = \mathbf{1}$ and $\psi_1 = \zeta_1$ of the Markov chain. (Remark the difference between $\bar{S}_n(\zeta_1, \mathbf{1})$ and the sum S_n in (2.16).) Lemma 6 implies that

$$|\bar{S}_n(\zeta_1, \mathbf{1}) - \bar{S}_n(\psi_1, x_1)| \le C_3, \tag{2.35}$$

where the constant C_3 depends only on the parameter ε from Condition **C**. But then, according to the law of large numbers applied to the Markov chain $(\omega_n, \zeta_n, y_n) \equiv (g_n, \zeta_n, y_n)$ defined in Theorem 4 we have that the following limit exists with probability 1:

$$\lim_{n\to\infty} \frac{1}{n} \log ||A_n \dots A_1|| = \lim_{n\to\infty} \frac{1}{n} \bar{S}_n(\zeta_1, y_1) = \lambda,$$

where λ is given by (2.13).

Formula (2.13) implies that the mean value of the function $f(g, z)$ defined by (2.19) is 0. Also, it is obvious that this function is Lipschitz on $\mathcal{J}_0 \times \mathrm{M}$ in all variables. Hence, Theorem 12 applies to the sequence S_n and statements (i), (ii), and (iii) of Theorem 6 are thus proved.

The case $\sigma = 0$ and $\lambda = 0$. Derivation of the algebraic condition for (P, Q, R). We start with a statement which is a corollary of a very general property proved in Lemma 13 from the Appendix.

Lemma 8. *Suppose that Condition* **C** *is satisfied and let $g \in \mathcal{J}_0$, $z_g \in \mathrm{M}$ be such that $g.z_g = z_g$. Then $z_g \in \mathrm{M}_0 \equiv \operatorname{supp}\nu$.*

Proof. According to Lemma 6, Condition **C** implies that every $g \in \mathcal{J}_0$ is contracting. Hence, by Lemma 13, $z_g \in \mathrm{M}_0$. □

Derivation of the algebraic condition. According to Theorem 12 (see formula (4.10)), the equality $\sigma = 0$ implies that $f(g, z) = F(z) - F(g.z)$. Hence, if z can be chosen to be equal to z_g, then it follows that $f(g, z_g) = 0$.

In the context of the present theorem the function f is given by $f(g, z) = \log ||(I - R - Q\psi)^{-1}Qx||$, where $g = (P, Q, R) \in \mathcal{J}_0$ and $z = (\psi, x) \in \mathrm{M}_0 \subset \Psi \times \mathbb{X}$. The equation $g.z_g = z_g$ is equivalent to saying that $z_g = (\psi, x)$ satisfies

$$(I - R - Q\psi)^{-1}\psi = \psi \text{ and } ||(I - R - Q\psi)^{-1}Qx||^{-1}(I - R - Q\psi)^{-1}Qx = x.$$

The equation $f(g, z_g) = 0$ now reads $\log ||(I - R - Q\psi)^{-1}Qx|| = 0$ or, equivalently, $||(I - R - Q\psi)^{-1}Qx|| = 1$. Hence the conditions $\sigma = 0$ and $\lambda = 0$ imply that all pairs $(g, z_g) \in \mathcal{J}_0 \times \mathrm{M}_0$ satisfy

$$(I - R - Q\psi)^{-1}P = \psi \text{ and } (I - R - Q\psi)^{-1}Qx = x.$$

But, by Lemma 5, this implies that $\mathcal{J}_0 \subset \mathcal{J}_{al}$, where $\mathcal{J}_{al}$ is defined by (1.8). $\square$

3. Proof of Theorem 1

As we are in the recurrent situation, we have that the Lyapunov exponent $\lambda = 0$.

Throughout this section we denote by C a generic positive constant which depends on nothing but ε and m and which may vary from place to place. If $f, g > 0$ are two functions, depending on $n \in \mathbb{Z},\ i \in \{1, \ldots, m\}$, and maybe on other parameters, we write

$$f \asymp g \text{ if there exists a } C > 1 \text{ such that } C^{-1}f \le g \le Cf.$$

Potential and its properties. As before, S_n is defined by (2.16). We put

$$\Phi_n(\omega) \equiv \Phi_n \stackrel{\text{def}}{=} \begin{cases} \log ||A_n \ldots A_1|| & \text{if } n \ge 1 \\ 0 & \text{if } n = 0 \\ -\log ||A_0 \ldots A_{n+1}|| & \text{if } n \le -1 \end{cases}, \tag{3.1}$$

where the matrices A_n are defined in (2.10). If $n \ge 1$, then obviously $\Phi_n \equiv \bar{S}_n(\zeta_1, \mathbf{1})$ defined in (2.34). The random function Φ_n is the analog of the potential considered first in [18]. For $n \ge a$, $a \in \mathbb{Z}$, put

$$S_{a,n}(\omega; \psi_a, x_a) \equiv S_{a,n}(\omega) \stackrel{\text{def}}{=} \log \|B_n \ldots B_a x_a\|, \tag{3.2}$$

where the matrices B_n are defined by (2.7). Similarly to (2.35), one has that

$$\left|S_{a,n}(\omega; \zeta_a, \mathbf{1}) - S_{a,n}(\omega; \psi_a, x_a)\right| \le C, \tag{3.3}$$

which implies:

$$\left|S_{a,n}(\omega) - (\Phi_n(\omega) - \Phi_a(\omega))\right| \le C. \tag{3.4}$$

Since one of the conditions of Theorem 1 is $\mathcal{J}_0 \not\subset \mathcal{J}_{al}$, it follows from Theorem 6, part (iv) that Φ_n satisfies the invariance principle with a strictly positive parameter $\sigma : \sigma > 0$.

The importance of the potential $\{\Phi_n\}_{n \in \mathbb{Z}}$ is due to that fact that it governs the stationary measure of our Markov chain; in fact it defines this stationary measure up to a multiplication by a bounded function (see (3.7). Namely, if $a < b$, we consider the Markov chain $\left\{\xi_t^{a,b}\right\}_{t\in\mathbb{N}}$ on

$$\mathbb{S}_{a,b} \stackrel{\text{def}}{=} \{a,\dots,b\} \times \{1,\dots,m\} \tag{3.5}$$

with transition probabilities (1.4) and reflecting boundary conditions at L_a and L_b. This means that we replace (P_a, Q_a, R_a) by $(I, 0, 0)$ and (P_b, Q_b, R_b) by $(0, I, 0)$. This reflecting chain has a unique stationary probability measure which we denote by $\pi_{a,b} = \left(\pi_{a,b}(k,i)\right)_{(k,i)\in\mathbb{S}_{a,b}}$. A description of this measure was given in [1]. We repeat it here for the convenience of the reader. To this end introduce row vectors $\nu_k \stackrel{\text{def}}{=} Z\left(\pi_{a,b}(k,i)\right)_{1\le i\le m}$, $a \le k \le b$, and Z is a (normalizing) factor. In terms of these vectors the invariant measure equation reads

$$\begin{aligned} \nu_k &= \nu_{k-1}P_{k-1} + \nu_k R_k + \nu_{k+1}Q_{k+1}, \quad \text{if } a < k < b, \\ \nu_a &= \nu_{a+1}Q_{a+1}, \quad \nu_b = \nu_{b-1}P_{b-1}. \end{aligned} \tag{3.6}$$

To solve Eq. (3.6), define for $a \le k < b$ matrices α_k by

$$\alpha_a \stackrel{\text{def}}{=} Q_{a+1}, \quad \text{and} \quad \alpha_k \stackrel{\text{def}}{=} Q_{k+1}\left(I - R_k - Q_k\psi_k\right)^{-1}, \quad \text{when } a < k < b,$$

where $\{\psi_k\}_{k\ge a+1}$ are given by (2.6) with the initial condition $\psi_{a+1} = I$ (we take into account that $R_a = Q_a = 0$ in our case). We shall now check that ν_k can be found recursively as follows: $\nu_k{=}\nu_{k+1}\alpha_k$, $a \le k < b$,, where ν_b satisfies $\nu_b\psi_b = \nu_b$. Indeed, the boundary condition at b in (3.6) reduces to $\nu_b = \nu_b\alpha_{b-1}P_{b-1} = \nu_b\psi_b$, where we use the fact that $\alpha_{b-1}P_{b-1} = \psi_b$ because $Q_b = I$ (and also due to (2.6)). But ψ_b is an irreducible stochastic matrix and therefore $\nu_b > 0$ exists and is uniquely defined up to a multiplication by a constant. We now have for $a < k < b$ that

$$\begin{aligned} \nu_{k-1}P_{k-1} + \nu_k R_k + \nu_{k+1}Q_{k+1} &= \nu_{k+1}\left(\alpha_k\alpha_{k-1}P_{k-1} + \alpha_k R_k + Q_{k+1}\right) \\ &= \nu_{k+1}\alpha_k\left(Q_k\psi_k + R_k + (I - R_k - Q_k\psi_k)\right) \\ &= \nu_{k+1}\alpha_k = \nu_k. \end{aligned}$$

Finally $\nu_a = \nu_{a+1}Q_{a+1}$ with $\alpha_a = Q_{a+1}$ and this finishes the proof of our statement.

We now have that

$$\pi_{a,b}(k,\cdot) = \pi_{a,b}(b,\cdot)\,\alpha_{b-1}\alpha_{b-2}\cdot\cdots\cdot\alpha_k,$$

where as before $\pi_{a,b}(k,\cdot)$ is a row vector. Note next that

$$\alpha_{b-1}\alpha_{b-2}\cdot\cdots\cdot\alpha_k = B_{b-1}\cdot\cdots\cdot B_{k+1}\left(I - R_k - Q_k\psi_k\right)^{-1}.$$

From this, we get

$$\pi_{a,b}(k,\cdot) \asymp \|B_{b-1}\cdot\cdots\cdot B_{k+1}\|\,\pi_{a,b}(b,\cdot),$$

and using (3.2), (3.4), we obtain for $a \le k, l \le b$,

$$\frac{\pi_{a,b}(k,\cdot)}{\pi_{a,b}(l,\cdot)} \asymp \exp\left[\Phi_k - \Phi_l\right]. \tag{3.7}$$

We also consider the "mirror situation" by defining for $n \leq a$ the matrices ψ_n^- in a similar way as in (2.6) by setting

$$\psi_{n-1}^- = \left(I - R_n - P_n \psi_n^-\right)^{-1} Q_n,\ n \leq a,$$

and a boundary condition ψ_a^-. Then, as in Theorem 4 a), one has that $\zeta_n^- \stackrel{\text{def}}{=} \lim_{a\to\infty} \psi_n^-$ exists almost surely, and does not depend on the boundary condition ψ_a^-. We then put

$$A_n^- \stackrel{\text{def}}{=} \left(I - R_n - P_n \zeta_n^-\right)^{-1} P_n,$$

and the potential Φ_n^- as (3.1):

$$\Phi_n^- \stackrel{\text{def}}{=} \begin{cases} \log \|A_0^- \dots A_{n-1}^-\| & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \\ -\log \|A_n^- \dots A_{-1}^-\| & \text{if } n \leq -1 \end{cases}.$$

We could as well have worked with this potential, and therefore we obtain

$$\frac{\pi_{a,b}(k,\cdot)}{\pi_{a,b}(l,\cdot)} \asymp \exp\left[\Phi_k^- - \Phi_l^-\right].$$

As $\Phi_0 = \Phi_0^- = 0$, we get

$$\left|\Phi_n - \Phi_n^-\right| \leq C \tag{3.8}$$

uniformly in n.

It is convenient to slightly reformulate the invariance principle for the potential. For that consider $C_0(-\infty,\infty)$, the space of continuous functions $f : (-\infty,\infty) \to \mathbb{R}$ satisfying $f(0) = 0$. We equip $C_0(-\infty,\infty)$ with a metric for uniform convergence on compacta, e.g.

$$d(f,g) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} 2^{-k} \min\left[1, \sup_{x\in[-k,k]} |f(x) - g(x)|\right], \tag{3.9}$$

and write $\mathcal{B}$ for the Borel-σ-field which is also the σ-field generated by the evaluation mappings $C_0(-\infty,\infty) \to \mathbb{R}$. We also write P_W for the law of the double-sided Wiener measure on $C_0(-\infty,\infty)$.

For $n \in \mathbb{N}$, we define

$$W_n\left(\frac{[k\sigma^2]}{n}\right) \stackrel{\text{def}}{=} \frac{\Phi_k}{\sqrt{n}},\ k \in \mathbb{Z},$$

and define $W_n(t)$, $t \in \mathbb{R}$, by linear interpolation. W_n is a random variable taking values in $C_0(-\infty,\infty)$.

Weak convergence of $\{W_n(t)\}_{t\in\mathbb{R}}$ on $C_0(-\infty,\infty)$ is the same as weak convergence of $\{W_n(t)\}_{t\in[-N,N]}$ for any $N \in \mathbb{N}$, and therefore, we immediately get

Proposition 7. *W_n converges in law to P_W.*

Let V be the subset of functions $f \in C_0(-\infty,\infty)$ for which there exist real numbers $a < b < c$ satisfying

1.

$$0 \in (a, c).$$

2.

$$f(a) - f(b) = f(c) - f(b) = 1.$$

3.

$$f(a) > f(x) > f(b), \quad \forall x \in (a, b),$$
$$f(c) > f(x) > f(b), \quad \forall x \in (b, c).$$

4. For any $\gamma > 0$,

$$\sup_{x \in (a-\gamma, a)} f(x) > f(a),$$
$$\sup_{x \in (c, c+\gamma)} f(x) > f(c).$$

It is clear that for $f \in V$, a, b, c are uniquely defined by f, and we write occasionally $a(f)$, $b(f)$, $c(f)$. $f(b)$ is the unique minimum of f in $[a, c]$. It is easy to prove that $V \in \mathcal{B}$, and

$$P_W(V) = 1.$$

If $\delta > 0$ and $f \in V$, we define

$$c_\delta(f) \stackrel{\text{def}}{=} \inf\{x > c : f(x) = f(c) + \delta\},$$
$$a_\delta(f) \stackrel{\text{def}}{=} \sup\{x < a : f(x) = f(a) + \delta\}.$$

If $\gamma > 0$, we set $V_{\delta,\gamma}$ to be the set of functions $f \in V$ such that

1.

$$c_\delta(f) \le 1/\delta, \; a_\delta(f) \ge -1/\delta. \tag{3.10}$$

2.

$$\sup_{b \le x < y \le c_\delta} [f(x) - f(y)] \le 1 - \delta, \tag{3.11}$$
$$\sup_{a_\delta \le y < x \le b} [f(x) - f(y)] \le 1 - \delta. \tag{3.12}$$

3.

$$\inf_{x \in [a_\delta, c_\delta] \setminus (b-\gamma, b+\gamma)} f(x) \ge f(b) + \delta. \tag{3.13}$$

It is evident that for any $\gamma > 0$, we have $V_{\delta,\gamma} \uparrow V$ for $\delta \downarrow 0$, and therefore, for any $\delta, \eta > 0$ we can find $\delta_0(\gamma, \eta)$ such that for $\delta \leq \delta_0$,

$$P_W\left(V_{\delta,\gamma}\right) \geq 1 - \eta.$$

It is easy to see that

$$P_W\left(\partial V_{\delta,\gamma}\right) = 0,$$

where ∂ refers to the boundary in $C_0(-\infty, \infty)$. Therefore, given $\gamma, \eta > 0$, we can find $N_0(\gamma, \eta)$ such that for $n \geq N_0$, $\delta \leq \delta_0$, we have

$$\mathbb{P}\left(W_n \in V_{\delta,\gamma}\right) \geq 1 - 2\eta. \tag{3.14}$$

For $t \in \mathbb{N}$, we set $n = n(t) \stackrel{\text{def}}{=} \left[\log^2 t\right]$. If $W_{n(t)} \in V_{\delta,\gamma}$, then we put

$$b_t \stackrel{\text{def}}{=} \frac{b\left(W_{n(t)}\right)\log^2 t}{\sigma^2},\ a_t \stackrel{\text{def}}{=} \frac{a_\delta\left(W_{n(t)}\right)\log^2 t}{\sigma^2},\ c_t \stackrel{\text{def}}{=} \frac{c_\delta\left(W_{n(t)}\right)\log^2 t}{\sigma^2}.$$

Remark that on $\left\{W_{n(t)} \in V_{\delta,\gamma}\right\}$, we have the following properties, translated from (3.10)-(3.13):

$$c_t \leq \tfrac{\log^2 t}{\sigma^2\delta},\ a_t \geq -\tfrac{\log^2 t}{\sigma^2\delta}, \tag{3.15}$$

$$\Phi_s - \Phi_{s'} \leq (1-\delta)\log t,\ b_t \leq s < s' \leq c_t, \tag{3.16}$$

$$\Phi_s - \Phi_{s'} \leq (1-\delta)\log t,\ a_t \leq s' < s \leq b_t, \tag{3.17}$$

$$\Phi_s \geq \Phi_{b_t} + \delta\log t,\ s \in [a_t, c_t] \setminus \left[b_t - \gamma\log^2 t, b_t + \gamma\log^2 t\right], \tag{3.18}$$

$$\min\left(\Phi_{a_t}, \Phi_{c_t}\right) - \Phi_{b_t} \geq (1+\delta)\log t. \tag{3.19}$$

Furthermore, if $0 \in [a_t, b_t]$, then

$$\sup_{0 \leq s \leq b_t} \Phi_s - \Phi_{b_t} \leq \log t, \tag{3.20}$$

and similarly if $0 \in [b_t, c_t]$.

(We neglect the trivial issue that a_t, b_t, c_t may not be in $\mathbb{Z}$.) The main result is

Proposition 8. *For $\omega \in \left\{W_{n(t)} \in V_{\delta,\gamma}\right\}$, we have for any $i \in \{1, \ldots, m\}$,*

$$Pr_{\omega,(0,i)}\left(X(t) \notin \left[b_t - \gamma\log^2 t, b_t + \gamma\log^2 t\right]\right) \leq 4t^{-\delta/2},$$

if t is large enough.

Together with (3.14), this proves our main result Theorem 1.

In all that follows, we keep γ, δ fixed, and assume that $\omega \in \left\{W_{n(t)} \in V_{\delta,\gamma}\right\}$. We will also suppress ω in the notation, and will take t large enough, according to ensuing necessities.

We first prove several estimates of probabilities characterizing the behaviour of a RW in a finite box in terms of the properties of the function S_n.

Lemma 9. *Consider a random walk on $\mathbb{S}_{a,b}$ with reflecting boundary conditions (see the discussion around (3.5)), and let $a < k < b$. Then*

$$Pr_{(k,i)}(\tau_a < \tau_b) \le C \sum_{y=k}^{b} \exp\left(\Phi_y - \Phi_a\right), \tag{3.21}$$

$$Pr_{(k,i)}(\tau_b < \tau_a) \le C \sum_{y=a}^{k} \exp\left(\Phi_y - \Phi_a\right). \tag{3.22}$$

Here τ_a, τ_b are the hitting times of the layers L_a, L_b.

Proof. We only have to prove (3.21). Equation (3.22) then follows in the mirrored situation and using (3.8).

Put $h_k(i) = Pr_{(k,i)}(\tau_b < \tau_a)$ and consider column-vectors $\mathbf{h}_k \stackrel{\text{def}}{=} (h_k(i))_{1\le i\le m}$. In order to find $\mathbf{h}_k$ we introduce the matrices $\varphi_{k+1} \stackrel{\text{def}}{=} (\varphi_{k+1}(i,j))_{1\le i,j\le m}$, where

$$\varphi_{k+1}(i,j) \stackrel{\text{def}}{=} Pr_{\omega,(k,i)}(\tau_{k+1} < \tau_a,\ \xi(\tau_{k+1}) = (k+1,j)). \tag{3.23}$$

These matrices satisfy (2.24) (with $a = 0$) with the modified boundary condition $\varphi_{a+1} = 0$. Equation (2.29) with ψ_k's defined by (2.6) now yields $\Delta_{k+1} = B_k \ldots B_{a+1}\psi_{a+1}\varphi_{a+2}\ldots\varphi_{k+1}$, and hence

$$\|\Delta_{k+1}\| \le \|B_k \ldots B_a\| \le C\exp(\Phi_k - \Phi_a). \tag{3.24}$$

The Markov property also implies that $\mathbf{h}_k = \varphi_{k+1}\mathbf{h}_{k+1}$, and hence

$$\mathbf{h}_k = \varphi_{k+1}\varphi_{k+2}\ldots\varphi_b\mathbf{1} \quad \text{since} \quad \mathbf{h}_b = \mathbf{1}. \tag{3.25}$$

We view the probabilities $Pr_{(k,\cdot)}(\tau_a < \tau_b)$ as the column vector $\mathbf{1}-\mathbf{h}_k$. Then, presenting $\varphi_b = \psi_b - \Delta_b$, we can have

$$\begin{aligned} Pr_{(k,\cdot)}(\tau_a < \tau_b) &= \mathbf{1} - \varphi_k \ldots \varphi_{b-1}\mathbf{1} = \mathbf{1} - \varphi_{k+1}\ldots\varphi_{b-1}(\psi_b - \Delta_b)\mathbf{1} \\ &= \mathbf{1} - \varphi_{k+1}\ldots\varphi_{b-1}\mathbf{1} + \varphi_{k+1}\ldots\varphi_{b-1}\Delta_b\mathbf{1} \\ &\le \mathbf{1} - \varphi_{k+1}\ldots\varphi_{b-1}\mathbf{1} + ||\Delta_b||\mathbf{1}. \end{aligned}$$

Iterating this inequality, we obtain that

$$Pr_{(k,\cdot)}(\tau_a < \tau_b) \le \sum_{y=k+1}^{b} ||\Delta_y||\mathbf{1}$$

and (3.21) follows from (3.24). □

Lemma 10. *Let $a < b$, and τ be the hitting time of $L_a \cup L_b$ – the union of two layers. Then if $a \le k \le b$, we have*

$$E_{(k,i)}(\tau) \le C(b-a)^2 \exp\left[\min\left(\sup_{a\le s<t\le b}(\Phi(s) - \Phi(t)),\ \sup_{a\le s<t\le b}(\Phi(t) - \Phi(s))\right)\right].$$

Proof. To prove that, consider column-vectors $\mathbf{e}_k = \left(E_{(k,i)}\tau\right)_{1\le i\le m}$. These vectors satisfy $\mathbf{e}_a = \mathbf{e}_b = \mathbf{0}$, and for $a < k < b$:

$$\mathbf{e}_k = P_k\mathbf{e}_{k+1} + R_k\mathbf{e}_k + Q_k\mathbf{e}_{k-1} + \mathbf{1}. \tag{3.26}$$

To solve (3.26), we use an induction procedure which allows us to find a sequence of matrices φ_k and vectors $\mathbf{d}_k$ such that

$$\mathbf{e}_k = \varphi_{k+1}\mathbf{e}_{k+1} + \mathbf{d}_k. \tag{3.27}$$

Namely, we put $\varphi_{a+1} = 0$, $\mathbf{d}_a = \mathbf{0}$ which according to (3.27) implies that $\mathbf{e}_a = \mathbf{0}$. Suppose next that φ_k and $\mathbf{d}_{k-1}$ are defined for some $k > a+1$. Then substituting $\mathbf{e}_{k-1} = \varphi_k\mathbf{e}_k + \mathbf{d}_{k-1}$ into the main equation in (3.26) we have

$$\mathbf{e}_k = P_k\mathbf{e}_{k+1} + R_k\mathbf{e}_k + Q_k(\varphi_k\mathbf{e}_k + \mathbf{d}_{k-1}) + \mathbf{1},$$

and hence

$$\mathbf{e}_k = (I - Q_k\varphi_k - R_k)^{-1}\left(P_k\mathbf{e}_{k+1} + Q_k\mathbf{d}_{k-1} + \mathbf{1}\right)$$

which makes it natural to put

$$\varphi_{k+1} = (I - Q_k\varphi_k - R_k)^{-1}P_k \tag{3.28}$$

and

$$\mathbf{d}_k = B_k(\varphi_k)\mathbf{d}_{k-1} + \mathbf{u}_k, \tag{3.29}$$

where

$$\mathbf{u}_k = (I - Q_k\varphi_k - R_k)^{-1}\mathbf{1}, \quad B_k(\varphi_k) = (I - Q_k\varphi_k - R_k)^{-1}Q_k.$$

The existence of matrices φ_k follows from the fact that $\varphi_k \ge 0$ and $\varphi_k\mathbf{1} \le \mathbf{1}$.

Iterating (3.27) and (3.29) we obtain

$$\mathbf{e}_k = \mathbf{d}_k + \varphi_{k+1}\mathbf{d}_{k+1} + \cdots + \varphi_{k+1}\ldots\varphi_{b-1}\mathbf{d}_{b-1}$$

and

$$\mathbf{d}_k = \mathbf{u}_k + B_k(\varphi_k)\mathbf{u}_{k-1} + \cdots + B_k(\varphi_k)\ldots B_{a+1}(\varphi_{a+1})\mathbf{u}_a.$$

Hence

$$\|\mathbf{e}_k\| \le \|\mathbf{d}_k\| + \|\mathbf{d}_{k+1}\| + \cdots + \|\mathbf{d}_{b-1}\| \le C(b-k)\max_{k\le j\le b-1}||\mathbf{d}_j||.$$

But $||B_k(\varphi_k)\ldots B_l(\varphi_l)|| \le C\sup_{a\le s<t\le b}\exp\left(\Phi(s) - \Phi(t)\right)$, and therefore

$$E_{(k,i)}(\tau) \le C(b-a)^2\exp\left[\sup_{a\le s<t\le b}\left(\Phi(s) - \Phi(t)\right)\right].$$

We obtain the same estimate with Φ replaced by Φ^-, and using (3.8), we get the desired estimate. □

Lemma 11. *Let $a \le k < b$ and $\xi(t)$ be as in Lemma 9. Then for any $x > 0$,*

$$Pr_{(k,i)}(\tau_b \ge x,\ \tau_b < \tau_a) \le \frac{C(b-a)^2}{x} \exp\left[\sup_{a \le s < t \le b}(\Phi(t) - \Phi(s))\right].$$

Proof. Let again τ being the hitting time of $L_a \cup L_b$. It is obvious that

$$Pr_{(k,i)}(\tau_b \ge x,\ \tau_b < \tau_a) \le Pr_{(k,i)}(\tau \ge x).$$

By the Markov inequality and Lemma 10, the result follows. □

Lemma 12. *Let $a < b$, and consider the chain $\{\xi_t\}$ on $\mathbb{S}_{a,b}$ with reflecting boundary conditions on a, b, as above. Then for any $t \in \mathbb{N}$, $(k,i),\ (l,j) \in \mathbb{S}_{a,b}$, we have*

$$Pr_{(k,i)}(\xi_t = (l,j)) \le C \exp[\Phi_l - \Phi_k].$$

Proof.

$$\begin{aligned}\pi_{a,b}(l,j) &= \sum_{(k',i')} \pi_{a,b}(k',i')\, Pr_{(k',i')}(\xi_t = (l,j)) \\ &\ge \pi_{a,b}(k,i)\, Pr_{(k,i)}(\xi_t = (l,j))\end{aligned}$$

for all $(k,i),\ (l,j) \in \mathbb{S}_{a,b}$, and all $t \in \mathbb{N}$. The lemma now follows with (3.7). □

We have now all the ingredients for the

Proof of Proposition 8. We may assume that $0 \in (a_t, b_t]$. The case of $0 \in (b_t, c_t)$ is handled similarly. We will write a, b, c for a_t, b_t, c_t, to simplify notations. We write J_t for the interval $\left[b - \gamma \log^2 t, b + \gamma \log^2 t\right]$.

We have

$$\begin{aligned}Pr_{(0,i)}(X(t) \notin J_t) \le{}& Pr_{(0,i)}(X(t) \notin J_t, \tau_b < \min(\tau_a, t)) + Pr_{(0,i)}(\tau_b > \tau_a) \\ &+ Pr_{(0,i)}(\tau_b > t, \tau_a > \tau_b).\end{aligned} \tag{3.30}$$

First we see that from Lemma 9, and (3.15), (3.19), (3.20),

$$\begin{aligned}Pr_{(0,i)}(\tau_b > \tau_a) &\le C(b-a)\exp\left[\sup_{0 \le x \le b} \Phi_x - \Phi_a\right] \\ &\le \frac{C \log^2 t}{\sigma^2 \delta} \exp\left[-\delta \log t\right] \le t^{-\delta/2},\end{aligned} \tag{3.31}$$

if t is large enough, and from Lemma 11 and (3.17),

$$\begin{aligned}Pr_{(0,i)}(\tau_b > t, \tau_a > \tau_b) &\le \frac{C \log^4 t}{t} \exp\left[\sup_{a \le s < t \le b}(\Phi(t) - \Phi(s))\right] \\ &\le \frac{C \log^4 t}{t} \exp\left[(1-\delta)\log t\right] \le t^{-\delta/2}.\end{aligned} \tag{3.32}$$

By the Markov property, we get

$$Pr_{(0,i)}\left(X(t) \notin J_t, \tau_b < \min(\tau_a, t)\right) \le \max_{s \le t, 1 \le j \le m} Pr_{(b,j)}\left(X(s) \notin J_t\right). \tag{3.33}$$

Now

$$Pr_{(b,j)}\left(X(s) \notin J_t\right) \le Pr_{(b,j)}\left(\min(\tau_a, \tau_c) \le t\right) + Pr_{(b,j)}\left(X^{(a,c)}(s) \notin J_t\right), \tag{3.34}$$

where $X^{(a,c)}$ is the chain with reflecting boundary conditions at L_a and L_c. The second summand is estimated by Lemma 12 and (3.18), which give

$$Pr_{(b,j)}\left(X^{(a,c)}(s) \notin J_t\right) \le C \exp\left[\sup_{l \notin J_t} \Phi_l - \Phi_b\right] \le C t^{-\delta} \le t^{-\delta/2}. \tag{3.35}$$

To estimate the first summand in (3.34) we observe that by (3.19),

$$\begin{aligned} Pr_{(b-1,i)}(\tau_a < \tau_b) &\le C \exp\left[-\Phi_a\right]\left(\exp\left[\Phi_{b-1}\right] + \exp\left[\Phi_b\right]\right) \le C \exp\left[-(1+\delta)\log t\right] \\ &\le t^{-1-2\delta/3}, \end{aligned}$$

and similarly

$$Pr_{(b+1,i)}(\tau_c < \tau_b) \le t^{-1-2\delta/3}.$$

If, starting in (b, j), the chain reaches L_a or L_c in time t, there is at least one among the first $t/2$ of the excursions from L_b which reaches $L_a \cup L_c$. By the above estimates, each such excursion has at most probability $t^{-1-2\delta/3}$ to be "successful", and therefore

$$Pr_{(b,j)}\left(\min(\tau_a, \tau_c) \le t\right) \le 1 - \left(1 - t^{-1-2\delta/3}\right)^{t/2} \le t^{-\delta/2}. \tag{3.36}$$

Combining (3.30)–(3.36), we get

$$Pr_{(0,i)}\left(X(t) \notin J_t\right) \le 4t^{-\delta/2}.$$

This proves the claim.

4. Appendix

Most (if not all) of the results in this appendix are not new. The main reason for including them is that we want to present them in the form which is needed for our purpose; this is particularly relevant in the case of Markov chains generated by contracting transformations. We also hope that a more self-contained paper makes an easier reading.

4.1. The CLT and the invariance principle (IP) for stationary Markov chains. We first recall, in Subsect. 4.1.1, the classical results of B. M. Brown [2] about the CLT and the IP for martingales. We then explain in Subsect. 4.1.2 that the reduction of the proof of the CLT for Markov chains to the martingale case invented by Gordin and Lifshits [10] can be easily extended to obtain the IP for Markov chains. Finally, in Subsect. 4.1.3, we prove that the Gordin-Lifshits conditions are satisfied for a class of Markov chains generated by contracting transformations.

4.1.1. The CLT and the IP for martingales (by B. M. Brown [2]). Let $\{ S_n, \mathcal{F}_n \}$, $n = 1, 2, \ldots$ be a martingale on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Put $U_n = S_n - S_{n-1}$ with $S_0 = 0$. The expectation with respect to $\mathbb{P}$ is denoted by $\mathbb{E}$, and $\mathbb{E}_{j-1}$ stands for the conditional expectation $\mathbb{E}(\cdot \mid \mathcal{F}_{j-1})$. Let $\sigma_n^2 = \mathbb{E}_{n-1}(U_n^2)$, $V_n^2 = \sum_{j=1}^n \sigma_j^2$, and $s_n^2 = \mathbb{E}(V_n^2) = \mathbb{E}(S_n^2)$. The main assumption in [2] concerned with martingales is:

$$V_n^2 s_n^{-2} \to 1 \quad \text{in probability as } n \to \infty. \tag{4.1}$$

We says that the Lindeberg condition holds for the class of martingales satisfying (4.1) if for any $\varepsilon > 0$,

$$s_n^{-2} \sum_{j=1}^{n} \mathbb{E} U_j^2 I(|U_j| \geq \varepsilon s_n) \to 0 \quad \text{as} \quad n \to \infty, \tag{4.2}$$

where $I(\cdot)$ is a characteristic function of a set.

For $t \in [0, 1]$ define a sequence of piecewise linear random functions

$$\begin{gathered} u_n(t) = s_n^{-1} \left(S_k + U_{k+1}(t s_n^2 - s_k^2)(s_{k+1}^2 - s_k^2)^{-1} \right) \\ \text{if } s_k^2 \leq t s_n^2 \leq s_{k+1}^2, \; k = 0, 1, \ldots, n-1. \end{gathered} \tag{4.3}$$

The following two theorems from [2] describe the asymptotic behaviour of the sequences S_n and $u_n(\cdot)$.

Theorem 9. *If (4.1) and (4.2) hold, then S_n is asymptotically normal:*

$$\lim_{n\to\infty} \mathbb{P}\{s_n^{-1} S_n \leq x \} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy \tag{4.4}$$

for all x. Furthermore, all finite dimensional distributions of $u_n(t)$ converge weakly, as $n \to \infty$, to those of a standard Wiener process $W(t)$ on $0 \leq t \leq 1$ (that is $W(0) = 0$ and $\mathbb{E}W^2(1) = 1$).

Theorem 10. *Let $\{ C[0,1], \mathcal{B}, P_W \}$ be the probability space where $C[0,1]$ is the space of continuous functions with the* sup *norm topology, $\mathcal{B}$ being the Borel σ-algebra generated by open sets in $C[0,1]$, and P_W the Wiener measure. Let $\{\mathbb{P}_n\}$ be the sequence of probability measures on $\{ C[0,1], \mathcal{B} \}$ determined by the distribution of $\{ u_n(t), 0 \leq t \leq 1 \}$. Then if (4.1) and (4.2) hold, $\mathbb{P}_n \to P_W$ weakly as $n \to \infty$.*

4.1.2. The CLT and the IP for general Markov chains. In their famous work [10], Gordin and Lifshits reduced the proof of the CLT for Markov chains to that of martingales. They then applied the same approach to the proof of the invariance principle for Markov chains in [11]. We shall explain their method here for the sake of completeness.

Let z_k, $k = 1, 2, \ldots$, be a stationary ergodic Markov chain with a phase space $(\mathfrak{X}, \mathcal{A})$, transition kernel $K(z, dy)$, and initial distribution κ. Let $f : \mathfrak{X} \mapsto \mathbb{R}$ be a real valued function on $\mathfrak{X}$ such that $\mathbb{E}f(z) = 0$ and $\operatorname{Var} f(z) < \infty$ (all expectations are taken with respect to the measure κ). Let $L_2(\mathfrak{X}, \mathcal{A}, \kappa)$ be the natural Hilbert space associated with $\mathfrak{X}, \mathcal{A}, \kappa$. By $\mathbf{I}$ we denote the identity operator in this space, and by $\mathfrak{A}$ the transition operator of the Markov chain: $\mathfrak{A}F(z) \stackrel{\text{def}}{=} \int_{\mathfrak{X}} F(y) K(z, dy)$. Put

$$S_n = f(z_1) + \cdots + f(z_n) \text{ with the convention } S_0 = 0. \tag{4.5}$$

Theorem 11. *Let z_k be a Markov chain described above and suppose that the function f with $\mathbb{E}f = 0$ can be presented as $f = (\mathbf{I} - \mathfrak{A})F$, where $F \in L_2(\mathfrak{X}, \mathcal{A}, \kappa)$ and $\mathbb{E}F = 0$. Put $\sigma^2 = ||F||^2 - ||\mathfrak{A}F||^2 \equiv \mathbb{E}F^2 - \mathbb{E}(\mathfrak{A}F)^2$ and suppose that $\sigma > 0$. Then $\frac{S_n}{\sigma\sqrt{n}}$ converges in law towards the standard Gaussian distribution $N(0, 1)$ and the sequence S_n satisfies the invariance principle with parameter σ in the sense of the definition given in Sect. 2.4.*

Proof. Consider the identity which is due to Gordin ([9]) and was used by Gordin and Lifshits in [10]: $f(z_k) = U(z_k, z_{k+1}) + F(z_k) - F(z_{k+1})$, where $U(z_k, z_{k+1}) = F(z_{k+1}) - (\mathfrak{A}F)(z_k)$. This identity holds true because of the conditions imposed on f. Obviously, $\mathbb{E}\{U(z_k, z_{k+1}) \,|\, z_k, \ldots, z_1\} = 0$. Denote $U_{k+1} = U(z_k, z_{k+1})$. In these notations we can write

$$S_n = \hat{S}_n + F(z_1) - F(z_{n+1}), \quad \text{where } \hat{S}_n = \textstyle\sum_{k=1}^n U_k.$$

It is clear that if $\mathcal{F}_n$ is a σ-algebra generated by the variables $z_1, \ldots, z_n$, then the sequence $\hat{S}_n$, $n = 1, 2, ..$ is a martingale with respect to the filtration $\mathcal{F}_n$, $n = 1, 2, \ldots$. Let us check that all conditions required by Theorems 9 and 10 are satisfied. Indeed, $\sigma_j^2 = \mathbb{E}\{U_j^2 \,|\, z_j\} = (\mathfrak{A}F^2)(z_j) - [(\mathfrak{A}F)(z_j)]^2$ is a stationary sequence with $\mathbb{E}\sigma_j^2 = ||F||^2 - ||\mathfrak{A}F||^2 = \sigma^2$. Relation (4.1) takes the form

$$(n\sigma^2)^{-1} \sum_{j=1}^{n} \sigma_j^2 \to 1$$

and is satisfied with probability 1 because of the Birkhoff Ergodic Theorem. The Lindeberg condition (4.2) takes the form

$$\mathbb{E}U_1^2 I(|U_1| \geq \varepsilon n\sigma^2) \to 0 \quad \text{as } n \to \infty,$$

and is obviously satisfied. Finally, functions (4.3) are now given by

$$u_n(t) = n^{-\frac{1}{2}}\sigma^{-1}\left(S_k + (tn - k)U_{k+1}\right) \quad \text{if } \; k \leq tn \leq k+1, \; k = 0, 1, \ldots, n-1,$$

and hence for $k \leq tn \leq k+1$,

$$v_n(t) = u_n(t) + n^{-\frac{1}{2}}\sigma^{-1}\left(F(z_1) - F(z_{k+1}) + (tn - k)(F(z_k) - F(z_{k+1}))\right),$$

where $v_n(t)$ is as in (2.17). Since F is square integrable and z_n is a stationary sequence, it follows that $n^{-\frac{1}{2}} \max_{1 \leq k \leq n} |F(z_k)| \to 0$ with probability 1 as $n \to \infty$. Hence also the $\sup_{0 \leq t \leq 1} |v_n(t) - u_n(t)| \to 0$ as $n \to \infty$ with probability 1. All statements of our theorem follow now from Theorems 9 and 10. □

4.1.3. The CLT and the IP for Markov chains generated by contracting transformations. Consider the following setup:

$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space; the related expectation is denoted $\mathbb{E}$.
M is a compact metric space equipped with a distance $\rho(\cdot, \cdot)$.

$\mathfrak{B}$ is a semigroup of continuous Lipschitz transformations of M: for any $g \in \mathfrak{B}$ there is a constant l_g such that $\rho(g.y, g.y') \leq l_g \rho(y, y')$ for any $y, \; y' \in$ M. Here and in the sequel $g.y$ denotes the result of the action of $g \in \mathfrak{B}$ on $y \in \mathrm{M}$; this notation will be used most of the time but in some cases we may write $g(y)$ rather than $g.y$.

For any $g_1, g_2 \in \mathfrak{B}$ put $\bar{\rho}(g_1, g_2) \stackrel{\text{def}}{=} \sup_{y \in \mathrm{M}} \rho(g_1.y, g_2.y)$. Obviously, $\bar{\rho}(\cdot, \cdot)$ defines a distance on $\mathfrak{B}$. We can now consider a Borel sigma-algebra generated by the corresponding open subsets of $\mathfrak{B}$; this sigma-algebra will be denoted by $\mathfrak{S}$.

Consider a measurable mapping $g : \Omega \mapsto \mathfrak{B}, \ \omega \mapsto g^{\omega}$ and for a $B \in \mathfrak{S}$ put $\mu(B) \stackrel{\text{def}}{=} \mathbb{P}\{\omega : g^{\omega} \in B\}$. We say that g is a random transformation of M. Let $g_k \in \mathfrak{B}, \ k \geq 1$ be a sequence of independent copies of g. Without loss of generality we can assume that g_k are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Denote by $\mathfrak{g}^{(j)} \stackrel{\text{def}}{=} g_j \ldots g_1$ the product of random transformations $g_1, \ldots, g_j$ and let $\mu^{(j)}$ be the probability distribution of the product $\mathfrak{g}^{(j)}$. This measure on $\mathfrak{B}$ is often called the j^{th} convolution power of the measure μ and is denoted by $\mu^{(j)} = \mu^{*j} = \mu * \cdots * \mu$ (j times).

A sequence of random transformations g_k is said to be *contracting* if there are constants $C > 0$ and $c, \ 0 \leq c < 1$ such that for any $y, \ y' \in \mathrm{M}$ and any $n \geq 1$,

$$\int_{\mathfrak{B}} \rho(g.y, g.y')\mu^{(n)}(dg) \equiv \mathbb{E}\rho(g_n \ldots g_1.y, g_n \ldots g_1.y') \leq Cc^n. \tag{4.6}$$

Remarks. Perhaps it would be more natural to say that the contraction property holds if $\int_{\mathfrak{B}} \rho(g.y, g.y')\mu^{(n)}(dg) \leq Cc^n \rho(y, y')$. However, (4.6) is sufficient for our purposes and is what we check in our applications.

As usual, products of random transformations generate a Markov chain with a state space M. Namely, let $\nu \equiv \nu(dy)$ be a probability measure on M and let $y_1 \in \mathrm{M}$ be chosen randomly according to the distribution ν and independent of all g_j's. For $k \geq 1$ define $y_{k+1} \in \mathrm{M}$ by $y_{k+1} \stackrel{\text{def}}{=} g_k.y_k \equiv \mathfrak{g}^{(k)}.y_1$. The sequence of pairs $(g_k, y_k), \ k \geq 1$ forms a Markov chain with a phase space $\mathfrak{B} \times \mathrm{M}$; this chain will be denoted $(\mathbf{g}, \mathbf{y})$. Note that the $(\mathbf{y})$-component of this chain, the sequences $y_k, \ k \geq 1$, is itself a Markov chain with the phase space M. Since M is a compact space the chain $(\mathbf{y})$ has an invariant measure; we shall suppose from now on that ν is such a measure which, in turn, implies that $\mu(dg)\nu(dy)$ is an invariant measure of the chain $(\mathbf{g}, \mathbf{y})$. It is well known (and easy to see) that if g_k is a contracting sequence of random transformations then the Markov chain $(\mathbf{y})$ has a unique invariant measure.

Let $\mathcal{L}_2(\mathfrak{B} \times \mathrm{M})$ be the Hilbert space of $\mu \times \nu$ square integrable real valued functions and $\mathcal{C}(\mathfrak{B} \times \mathrm{M})$ be its subset of continuous functions.

Given an $f \in \mathcal{C}(\mathfrak{B} \times \mathrm{M})$ let S_n denote the related Birkhoff sums along a trajectory of the Markov chain $(\mathbf{g}, \mathbf{y})$:

$$S_n = \sum_{k=1}^{n} f(g_k, y_k).$$

By $\mathfrak{A}$ we denote the following Markov operator acting in $\mathcal{L}_2(\mathfrak{B} \times \mathrm{M})$ and preserving $\mathcal{C}(\mathfrak{B} \times \mathrm{M})$:

$$(\mathfrak{A}f)(g, y) \stackrel{\text{def}}{=} \int_{\mathfrak{B}} f(g', g.y)\mu(dg'). \tag{4.7}$$

It follows from (4.7) that

$$(\mathfrak{A}^k f)(g, y) = \int_{\mathfrak{B} \times \mathfrak{B}} f(g', \tilde{g}g.y)\mu(dg')\mu^{(k-1)}(d\tilde{g}). \tag{4.8}$$

Theorem 12. *Suppose that the sequence of random transformations g_k is contracting and f is a continuous bounded function on $\mathfrak{B} \times \mathrm{M}$ such that*

(i) $\int_{\mathfrak{B}} f(g, y)\mu(dg)$ is Lipschitz on M, *that is for some C_f*

$$\left|\int_{\mathfrak{B}} (f(g, y) - f(g, y'))\mu(dg)\right| \le C_f \rho(y, y').$$

(ii) $\int_{\mathfrak{B}} f(g, y)\mu(dg)\nu(dy) = 0$.

Then the equation

$$(I - \mathfrak{A})F = f, \tag{4.9}$$

has a solution $F(g, y)$ which is continuous on $\mathfrak{B} \times \mathrm{M}$ and

$$\int_{\mathfrak{B}\times \mathrm{M}} F(g, y)\mu(dg)\nu(dy) = 0.$$

Besides, this solution is unique in $\mathcal{L}_2(\mathfrak{B} \times \mathrm{M})$.

Denote by

$$\sigma^2 = \int_{\mathfrak{B}\times \mathrm{M}} (\mathfrak{A}F^2 - (\mathfrak{A}F)^2)(g, y)\mu(dg)\nu(dy).$$

If $\sigma > 0$ then $\frac{S_n}{\sigma\sqrt{n}}$ converges in law towards the standard Gaussian distribution $N(0, 1)$ and the sequence S_n satisfies the invariance principle with parameter σ. If $\sigma > 0$ and, in addition to (i), $|f(g, y) - f(g, y')| \le C_f(g)\rho(y, y')$ with $\int \log(1+C_f(g))\mu(dg) < \infty$, then the invariance principle for the sequence S_n is satisfied uniformly in $y_1 \in \mathrm{M}$.

If $\sigma = 0$, then the function $F(g, y)$ depends only on y and for every (g, y) in the support of $\mu \times \nu$ one has

$$f(g, y) = F(y) - F(g.y). \tag{4.10}$$

Proof. The existence of F. Equation (4.9) can be rewritten as $F = \mathfrak{A}F + f$ and, iterating this relation, one obtains a formal series:

$$F = \sum_{k=0}^{\infty} \mathfrak{A}^k f \tag{4.11}$$

Condition (ii) of the theorem and the invariance of the measure $\mu(dg)\nu(dy)$ imply that

$$\int_{\mathfrak{B}\times \mathrm{M}} (\mathfrak{A}^k f)(g, y)\mu(dg)\nu(dy) = \int_{\mathfrak{B}\times \mathrm{M}} f(g, y)\mu(dg)\nu(dy) = 0.$$

Hence, the convergence in (4.11) would follow if we prove that

$$|(\mathfrak{A}^k f)(g, y) - (\mathfrak{A}^k f)(\bar{g}, \bar{y})| \le \text{const}\, c^{\frac{k}{n_0}} \text{ for any } (g, y), (\bar{g}, \bar{y}) \in \text{support of } \mu \times \nu. \tag{4.12}$$

But it follows from (4.8) and condition (i) of the theorem that

$$\begin{aligned} &|(\mathfrak{A}^k f)(g, y) - (\mathfrak{A}^k f)(\bar{g}, \bar{y})| \\ &= \left|\int_{\mathfrak{B}} \left(\int_{\mathfrak{B}} \left(f(g', \tilde{g}g.y) - f(g', \tilde{g}\bar{g}.\bar{y})\right) \mu(dg')\right) \mu^{(k-1)}(d\tilde{g})\right| \\ &\le C_f \int_{\mathfrak{B}} \rho(\tilde{g}g.y, \tilde{g}\bar{g}.\bar{y})\mu^{(k-1)}(d\tilde{g}) \le C\, c^n, \end{aligned}$$

where the last inequality is due to the contraction property (4.6). The existence and continuity of $F(g, y)$ is proved.
Uniqueness. As usual, to prove the uniqueness we have to show that the homogeneous equation $F = \mathfrak{A}F$ has only a trivial solution $F \equiv 0$ in the class of functions satisfying the condition $\int_{\mathfrak{B}\times \mathrm{M}} F(g, y)\mu(dg)\nu(dy) = 0$. To check that this is the case assume that, to the contrary, there is an $F \in \mathcal{L}_2(\mathfrak{B} \times \mathrm{M})$ such that $F \not\equiv 0$, satisfies the homogeneous equation, and has a zero mean value. For a given $\epsilon > 0$ find a function $\tilde{F}$ which is Lipschitz on $\mathfrak{B} \times \mathrm{M}$ and approximates F in the sense that $||F - \tilde{F}|| \le \epsilon$, where $|| \cdot ||$ denotes the $\mathcal{L}_2(\mathfrak{B} \times \mathrm{M})$ norm. The $\tilde{F}$ can always be chosen so that $\int_{\mathfrak{B}\times \mathrm{M}} \tilde{F}(g, y)\mu(dg)\nu(dy) = 0$. Next, for any $n \ge 1$,

$$F = \mathfrak{A}^n F = \mathfrak{A}^n (F - \tilde{F}) + \mathfrak{A}^n \tilde{F}.$$

But then $\mathfrak{A}^n \tilde{F} \to 0$ uniformly in (g, y) and $||\mathfrak{A}^n (F - \tilde{F})|| \le \epsilon$. Since ϵ can be made arbitrarily small, we conclude that $F \equiv 0$.
Proof of the CLT and the IP in the case $\sigma > 0$. According to Theorem 11 the existence of $F \in \mathcal{L}_2(\mathfrak{B} \times \mathrm{M})$ satisfying Eq. (4.9) is the main condition under which both the Central Limit Theorem and the Invariance Principle hold for Birkhoff sums picked up along a realization of a trajectory of a Markov chain. The ergodicity of the Markov chain is the other condition which is needed and which in our case follows from the contraction property. The CLT and the IP is thus proved.
Proof of the uniform IP in the case $\sigma > 0$. We write $S_n(y_1)$ for S_n in order to emphasize the dependence of this sequence on y_1. Clearly,

$$|S_n(y_1) - S_n(y_1')| \le \sum_{k=1}^{n} |f(g_k, y_k) - f(g_k, y_k')| \le \sum_{k=1}^{\infty} C_f(g_k)\rho(y_k, y_k'). \tag{4.13}$$

It follows from (4.6) (due to the Chebyshev inequality) that $\mathbb{P}$ almost surely $\rho(y_k, y_k') \le e^{-\varepsilon k}$ for some $\varepsilon > 0$ and $k \ge k(\varepsilon, \omega)$. It is essential that $k(\varepsilon, \omega)$ does not depend on y_1, y_1'. Next, due to the condition imposed on the function f, the sequence $k^{-1}\log(1 + C_f(g_k)) \to 0$ as $k \to \infty$ $\mathbb{P}$ almost surely. Hence the right-hand side of (4.13) is $\mathbb{P}$ almost surely bounded and the corresponding estimate does not depend on y_1, y_1'.

Let us now consider the dependence on y_1 of the relevant $v_n(t) = v_n(t; y_1)$ (see (2.17)). For $t \in [0, 1]$, and $k \le tn \le k + 1,\ k = 0, 1, \ldots, n - 1$ we have:

$$v_n(t; y_1) - v_n(t; y_1') = n^{-\frac{1}{2}} \left(S_k(y_1) - S_k(y_1') + (f_{k+1}(y_1) - f_{k+1}(y_1'))(tn - k)\right)$$

with the obvious meaning of $f_{k+1}(y_1)$ and $f_{k+1}(y_1')$. It is now clear that $\mathbb{P}$ almost surely $v_n(t; y_1) - v_n(t; y_1') \to 0$ as $n \to \infty$ uniformly in y_1, y_1'. This proves that the uniformity of the invariance principle.
The case $\sigma = 0$. Note that

$$(\mathfrak{A}F^2 - \mathfrak{A}(F^2))(g, y) = \int_{\mathfrak{B}} \left(F(g', g.y) - \int_{\mathfrak{B}} F(\tilde{g}, g.y)\mu(d\tilde{g})\right)^2 \mu(dg').$$

Hence $\sigma = 0$ implies that for $\mu \times \nu$-almost all (g, y) and μ-almost all g'

$$F(g', g.y) = \int_{\mathfrak{B}} F(\tilde{g}, g.y)\mu(d\tilde{g}). \tag{4.14}$$

But $F(\cdot,\cdot)$ is a continuous function of both variables and hence (4.14) holds for any (g, y) from the support of $\mu \times \nu$. This proves that F depends only on the second variable: $F(g', g.y) \equiv F(g.y)$ (we note that $g.y$ runs over the whole of the support of ν when (g, y) runs over the support of $\mu \times \nu$). Finally, one obtains (4.10) by substituting $F(y)$ (rather than $F(g, y)$) into (4.9). □

4.1.4. Markov chains generated by contracting transformations: characterization of the support of the invariant measure. The aim of this section is to give a characterization of the support of an invariant measure of a Markov chain generated by contracting transformations in terms of fixed points of these transformations.

We work here within the same setup as in Sect. 4.1.3. This applies to the sequence g_j, $j \geq 1$, the metric space (M, ρ), the semigroup $\mathfrak{B}$ of transformations of M, the Markov chain y_j defined by $y_{j+1} = g_j.y_j$, $j \geq 1$ (with y_1 being a random element independent of all g_j's). However, we shall suppose that $\mathfrak{B}$ is generated by the transformations belonging to the support $\mathcal{J}_0$ of the distribution μ of g_j's. This difference is important for Lemma 14.

Let ν be the stationary measure of our chain and M_0 be the support of ν.

As usual, we say that a transformation $g \in \mathfrak{B}$ is a contraction on a subset $\mathrm{M}_0 \subset \mathrm{M}$ if there is an $n \geq 1$ and a $c \in [0, 1)$ (both n and c may depend on g) such that $\rho(g^n.x', g^n.x'') \leq c\rho(x', gx'')$ for any $x', x'' \in \mathrm{M}_0$. If $g \in \mathfrak{B}$, then by x_g we denote a fixed point of the transformation g: $g.x_g = x_g$.

Lemma 13. *If $g \in \mathfrak{B}$ is a contraction on* M *then its fixed point $x_g \in \mathrm{M}$, belongs to the support* M_0 *of the invariant measure ν of the Markov chain y_j.*

Proof. Consider a random infinite sequence $g_1, g_2, \ldots$. Since $g \in \mathcal{J}_0$, almost every such sequence has the property that for any $k \geq 1$ and any $\delta > 0$ there are infinitely many i's such that each element of the part $g_i, \ldots, g_{i+nk-1}$ of the sequence approximates g so closely that

$$\bar{\rho}(g^{nk}, \mathfrak{g}_i^{(nk)}) \leq \delta \quad \text{where} \quad \mathfrak{g}_i^{(nk)} \stackrel{\text{def}}{=} g_{i+nk-1} \cdots g_i.$$

Moreover, by the law of large numbers these i's have a positive frequency. Since

$$\rho(x_g, g^{nk}.x') = \rho(g^{nk}x_g, g^{nk}.x') \leq c^k \rho(x_g, x')$$

for any $x' \in \mathrm{M}$, we have that

$$\rho(x_g, \mathfrak{g}_i^{(nk)}.x') \leq c^k \rho(x_g, x') + \rho(g^{nk}.x', \mathfrak{g}_i^{(nk)}.x') \leq c^k \rho(x_g, x') + \delta.$$

Hence any (small) neighbourhood of x_g is visited by the sequence $\mathfrak{g}_1^{(j)}.x'$, $j \geq 1$, infinitely many times and, moreover, this happens with a positive frequency for almost every sequence g_j, $j \geq 1$. This implies that $x_g \in \mathrm{M}_0$ and $(g, x_g) \in \mathcal{J}_0 \times \mathrm{M}_0$. □

Note that if the invariant measure ν of our Markov chain is ergodic, then the support M_0 of this measure is a minimal set of $\mathfrak{B}$. The latter by definition means that the orbit $\{g.x : g \in \mathfrak{B}\}$ of any $x \in \mathrm{M}_0$ is everywhere dense in M_0.

Lemma 14. *Let $\mathrm{M}_0 \subset \mathrm{M}$ be a minimal set of $\mathfrak{B}$. Suppose that there exist a $\hat{g} \in \mathfrak{B}$ which is a contraction on M_0. Consider the set of all fixed points of $\mathfrak{B}$ belonging to M_0:*

$$\mathrm{Fix}_{\mathrm{M}_0}(\mathfrak{B}) \stackrel{\text{def}}{=} \{x : x \in \mathrm{M}_0 \text{ and there is a } g \in \mathfrak{B} \text{ such that } g.x{=}x\}.$$

Then $\mathrm{Fix}_{\mathrm{M}_0}(\mathfrak{B})$ is everywhere dense in M_0.

Proof. The contraction $\hat{g}$ given to us by the condition of the lemma has a fixed point $\hat{x} \in \mathrm{M}_0$ (it may have other fixed points too, but we are interested only in this one). Since M_0 is minimal it coincides with the closure of the orbit $\{g.\hat{x} \, : \, g \in \mathfrak{B}\}$. For a given $g \in \mathfrak{B}$ let us consider the point $g.\hat{x}$. We shall now show that for a sufficiently large n the transformation $g\hat{g}^n$ has a fixed point which we shall denote $x_{g\hat{g}^n}$. Indeed, for any $x', x'' \in \mathrm{M}_0$,

$$\rho(g\hat{g}^n.x', g\hat{g}^n.x'') \leq l_g \rho(\hat{g}^n.x', \hat{g}^n.x'') \leq l_g c^n \rho(x', x'').$$

If n is such that $l_g c^n < 1$, then there is a fixed point $x_{g\hat{g}^n}$ of $g\hat{g}^n$. On the other hand, it is obvious that $g\hat{g}^n.x' \to g.\hat{x}$ as $n \to \infty$ uniformly in $x' \in \mathrm{M}_0$ because $\hat{g}^n.x' \to \hat{x}$ uniformly in $x' \in \mathrm{M}_0$. It follows that in particular $x_{g\hat{g}^n} \to g.\hat{x}$ and this proves the lemma. $\square$

4.2. Products of positive matrices. Lemma 15 below explains two versions of a well known contraction property of products of positive matrices (see, e.g. [5]). The first version of this property has already been explained and proved in the Appendix to [1] and we therefore prove here only the second version. There is a slight difference in the notations used in this paper and those we have introduced in [1] and no difference in the proof; we emphasize once again that this is done for the purposes of completeness and convenience of references in the proofs of other theorems.

Lemma 15. *Let $a_n = (a_n(i, j))$, $n = 1, 2, \ldots$ be a sequence of positive $m \times m$ matrices, $a_n > 0$. Put $\tilde{H}_n \stackrel{\text{def}}{=} a_n a_{n-1} \ldots a_1$, $H_n \stackrel{\text{def}}{=} a_1 a_2 \ldots a_n$ and denote*

$$\tilde{\delta}_r = \min_{i,j,k} a_r(i, j)a_{r-1}(j, k) \left(\sum_j a_r(i, j)a_{r-1}(j, k) \right)^{-1}, \quad 2 \leq r \leq n,$$

$$\delta_r = \min_{i,j,k} a_r(i, j)a_{r+1}(j, k) \left(\sum_j a_r(i, j)a_{r+1}(j, k) \right)^{-1}, \quad 1 \leq r \leq n-1. \tag{4.15}$$

Suppose that

$$\sum_{r=2}^{\infty} \tilde{\delta}_r = \infty.$$

Then the products H_n and $\tilde{H}_n$ can be presented as follows:

$$H_n = D_n[(c_n(1)\mathbf{1}, \ldots, c_n(m)\mathbf{1}) + \phi_n], \quad \tilde{H}_n = \tilde{D}_n[(\tilde{c}(1)\mathbf{1}, \ldots, \tilde{c}(m)\mathbf{1}) + \tilde{\phi}_n], \tag{4.16}$$

where:

D_n and $\tilde{D}_n$ are diagonal matrices with positive diagonal elements;

$\|\phi_n\| \leq \prod_{r=1}^{n-1}(1 - m\delta_r)$, $\left\|\tilde{\phi}_n\right\| \leq \prod_{r=2}^{n}(1 - m\tilde{\delta}_r)$;

$\tilde{c}(j)$ are strictly positive numbers which are uniquely defined by the sequence $\{a_k\}_{k\geq1}$, do not depend on n, and such that $\sum_j \tilde{c}(j) = 1$;

$c_n(j)$ are strictly positive numbers with $\sum_j c_n(j) = 1$ (note that $c_n(j)$, unlike the $\tilde{c}(j)$, do depend on n and, generally, do not have a limit).

Proof. Present H_n as follows:

$$H_n = D_n D_n^{-1} a_1 D_{n-1} D_{n-1}^{-1} a_2 \dots D_1^{-1} a_n = D_n \tilde{a}_1 \tilde{a}_2 \dots \tilde{a}_n,$$

where $\tilde{a}_r \equiv D_{n-r+1}^{-1} a_r D_{n-r}$, $\tilde{D}_0 \stackrel{\text{def}}{=} I$, and $D_{n-r} = \operatorname{diag}(D_{n-r}(1), \dots, D_{n-r}(m))$ are diagonal matrices, with $D_{n-r}(i)$ chosen so that to make matrices $\tilde{a}_r$ stochastic. It is very easy to see that the only such choice is given by

$$D_{n-r}(i) = \sum_{i_{r+1},\dots,i_n} a_{r+1}(i, i_{r+1})\, a_{r+2}(i_{r+1}, i_{r+2}) \dots a_n(i_{n-1}, i_n)$$

and

$$\tilde{a}_r(i,j) = \frac{a_r(i,j) \sum_{i_{r+1},\dots,i_n} a_{r+1}(j, i_{r+1}) \dots a_n(i_{n-1}, i_n)}{\sum_{i_r, i_{r+1},\dots,i_n} a_r(i, i_r)\, a_{r+1}(i_r, i_{r+1}) \dots a_n(i_{n-1}, i_n)} \geq \delta_r. \tag{4.17}$$

It is well known that the last estimate implies the following presentation of the product of stochastic matrices $\tilde{a}_n$:

$$\tilde{a}_1 \tilde{a}_2 \dots \tilde{a}_n = (c_n(1)\mathbf{1}, \dots, c_n(m)\mathbf{1}) + \phi_n,$$

where

$$\min_i \tilde{a}_n(i,j) \leq c_n(j) \leq \max_i \tilde{a}_n(i,j) \tag{4.18}$$

and the matrices ϕ_n are such that

$$\|\phi_n\| \leq \prod_{r=1}^{n-1} (1 - m\delta_r).$$

□

4.3. A stability estimate. The stability property which we explain below is definitely well known to specialists in the relevant field. Given that the proof is very short, it seems that it is easier for us to prove it than to find a relevant reference.

Let b_n and b'_n be two sequences of transformations of a metric space $(\mathbb{X}, \mathfrak{r})$ and $x_{n+1} \stackrel{\text{def}}{=} b_n(x_n)$, $x'_{n+1} \stackrel{\text{def}}{=} b'_n(x'_n)$, $n \geq 1$, with given initial values $x_1,\ x'_1 \in \mathbb{X}$. For any two transformations b and b' put $\bar{\rho}(b, b') \stackrel{\text{def}}{=} \sup_{x \in \mathbb{X}} \mathfrak{r}(b(x), b'(x))$.

Lemma 16. *Suppose that*

(a) b_n are uniformly contracting, that is there is a c, $0 \leq c < 1$, such that for any $x, y \in \mathbb{X}$ we have $\mathfrak{r}(b_n(x), b_n(y)) \leq c\,\mathfrak{r}(x, y)$;

(b) $\bar{\rho}(b_n, b'_n) \to 0$ as $n \to \infty$.

Then $\mathfrak{r}(x_n, x'_n) \to 0$ as $n \to \infty$.

If, instead of (b), a stronger property holds, namely $\bar{\rho}(b_n, b'_n) \leq C_2 c_0^n \bar{\rho}(b_1, b'_1)$ for some C_2 and $c_0 < 1$, then for $\epsilon > 0$ there is a constant C_3 such that

$$\mathfrak{r}(x_n, x'_n) \leq C_3 \tilde{c}^n (\bar{\rho}(b_1, b'_1) + \mathfrak{r}(x_1, x'_1)), \text{ where } \tilde{c} = \max(c, c_0) + \epsilon. \tag{4.19}$$

Proof. Put $d_n \stackrel{\text{def}}{=} \bar{\rho}(b_n, b'_n)$ and $r_n \stackrel{\text{def}}{=} \mathfrak{r}(x_n, x'_n)$. Since

$$\begin{aligned}\mathfrak{r}(x_{n+1}, x'_{n+1}) &= \mathfrak{r}(b_n(x_n), b'_n(x'_n)) \le \mathfrak{r}(b_n(x_n), b_n(x'_n)) + \mathfrak{r}(b_n(x'_n), b'_n(x'_n)) \\ &\le c\mathfrak{r}(x_n, x'_n) + \bar{\rho}(b_n, b'_n),\end{aligned}$$

we have that

$$r_{n+1} \le cr_n + d_n \le d_n + cd_{n-1} + \cdots + c^k d_{n-k} + c^{k+1} r_{n-k}. \tag{4.20}$$

For a given $\epsilon > 0$ choose k so that $c^k r_{n-k} \le \epsilon$ (which is possible because $\mathbb{X}$ is a compact space and thus r_{n-k} is a uniformly bounded sequence). Next choose $N(\epsilon, k)$ so that $d_{n-j} \le \epsilon$ when $n - j \ge N(\epsilon, k) - k$. It follows now from (4.20) that $r_n \le (2 - c)(1 - c)^{-1}\epsilon$ when $n > N(\epsilon, k)$. This proves the first statement of the lemma.

To prove the second statement substitute $k = n$ into (4.20) and take into account the stronger estimates for d_n. Estimate (4.19) follows with an evident choice of C_3. □

Remarks. The second statement of this lemma does not use the fact that $\mathbb{X}$ is a compact space.

Acknowledgement. This work was supported by the following grants of the Swiss National Foundation: 200020-107739/1 and 200020-116348. We are grateful to the Isaac Newton Institute for its hospitality during the program *Interaction and Growth in Complex Stochastic Systems* held in Cambridge, UK in 2003. We also thank the European Science Foundation Research Networking Programme on *Phase-Transitions and Fluctuation Phenomena for Random Dynamics in Spatially Extended Systems (RDSES)* for its financial support.

References

1. Bolthausen, E., Goldsheid, I.: Recurrence and transience of random walks in random environments on a strip. Commun. Math. Phys. **214**, 429–447 (2000)
2. Brémont, J.: On some random walks on Z in random medium. Ann. Probab. **30**, 1266–1312 (2002)
3. Brémont, J.: Behavior of random walks on Z in Gibbsian medium. C. R. Acad. Sci. Série 1 Math. **338**(11), 895–898 (2004)
4. Brown, B.M.: Martingale Central Limit Theorems. Ann. Math. Statist. **42**, 59–66 (1971)
5. Furstenberg, H., Kesten, H.: Products of random matrices. Ann. Math. Statist. **31**, 457–469 (1960)
6. Goldsheid, I.: Linear and Sub-linear Growth and the CLT for Hitting Times of a Random Walk in Random Environment on a Strip. Probability Theory and Related Fields, appeared on line in August, 2007, DOI:10.1007/s00440-007-0091-0
7. Golosov, A.: Localization of random walks in one-dimensional random environments. Commun. Math. Phys. **92**, 491–506 (1984)
8. Golosov, A.: On the limit distributions for a random walk in a critical one-dimensional random environment. Usp. Mat. Nauk **41**(2), 189–190 (1986)
9. Gordin, M.I.: The Central Limit Theorem for stationary processes. Soviet Math. Dokl. **10**, 1174–1176 (1969)
10. Gordin, M.I., Lifshits, B.A.: The Central Limit Theorem for stationary Markov processes. Sov. Math. Dokl. **19**(2), 392–394 (1978)
11. Gordin, M.I., Lifshits, B.A.: The Invariance principle for stationary Markov processes. "Teorija verojatnostej i ejo primenenija" 1978, issue 4, pp. 865-866 (in Russian)
12. Hall, P., Heyde, C.C.: *Martingale limit theory and its application.* New York: Academic Press, 1980
13. Kesten, H.: The limit distribution of Sinai's random walk in a random environment. Physica A **138**, 299–309 (1986)
14. Kesten, H., Kozlov, M.V., Spitzer, F.: Limit law for random walk in a random environment. Comp. Math. **30**, 145–168 (1975)
15. Key, E.: Recurrence and transience criteria for random walk in a random environment. Ann. Prob. **12**, 529–560 (1984)
16. Lawler, G.: Weak convergence of a random walks in a random environment. Commun. Math. Phys. **87**, 81–87 (1982)

17. Letchikov, A.V.: *Localization of one-dimensional random walks in random environment. Soviet Scientific Reviews Section C: Mathematical Physics Reviews.* Chur. Switzerland: Harwood Academic Publishers, 1989, pp. 173–220
18. Sinai, Ya.G.: The limiting behavior of a one-dimensional random walk in a random medium. Theory Prob. Appl. **27**, 256–268 (1982)
19. Solomon, F.: Random walks in a random environment. Ann. Prob. **3**, 1–31 (1975)
20. Zeitouni, O.: *Random walks in random environment*, XXXI Summer school in Probability, St. Flour (2001). Lecture notes in Math. **1837**, Berlin:Springer, 2004, pp. 193–312

Communicated by M. Aizenman