Random Matrix Products and the Statistical Mechanics of Disordered Systems

Giambattista Giacomin

Version March 26, 2024

5045 Bât. Sophie Germain (Campus PRG)
Email address: giacomin@lpsm.paris
URL: https://www.lpsm.paris/users/giacomin/index

CHAPTER 1

Products of random 2×2 matrices

1.1. Foreword: products of IID random variables and matrices

The focus of this chapter is on the $n \to \infty$ asymptotic behavior of products of the form $Y_n Y_{n-1} \ldots Y_1$ or $Y_1 \ldots Y_{n-1} Y_n$ when $Y_1, Y_2, \ldots$ are IID real random matrices, and we will limit ourselves to two by two matrices. We stress that by this we do not mean that the entries of Y_1 are independent random variables: they can be, but we will actually consider very general joint laws of the entry of the matrices.

By *behavior of products* of matrices we may mean several things: for example behavior of the entries, behavior of a norm of the matrix product, behavior of the matrix product when applied to a given vector (again, we can take the norm, or look at the components). The first and, in a sense, most important case is considering a norm of the matrix (we will see that which norm we choose is irrelevant). In fact we will consider the limit of

$$\frac{1}{n} \log \|Y_n Y_{n-1} \ldots Y_1\| , \tag{1.1}$$

and we will show that, under suitable assumptions, the limit exists a.s. and in $\mathbb{L}^1$. In this case, the limit is not random and does not depend on the norm: we suggest to check the norm independence (use that norms on finite dimensional spaces are equivalent). The value of the limit is called *top Lyapunov exponent* and, in these notes, is denoted by γ. Of course, it just depends on the law of Y_1. We will see also that the limit is the same if we consider the product $Y_1 Y_2 \ldots Y_n$ instead.

Before looking at the case of matrices, it is natural to quickly review the case of scalars: if $Y_1, Y_2, \ldots$ are IID real random real numbers (variables). In this case Kolmogorov's Law of Large Numbers tells us that

$$\lim_{n\to\infty} \frac{1}{n} \log |Y_1 Y_2 \ldots Y_n| \, = \, \mathbb{E}\left[\log |Y_1|\right] , \tag{1.2}$$

a.s. if and only if $\mathbb{E}\left[|\log |Y_1||\right] < \infty$. The convergence holds also in $\mathbb{L}^1$. From this result, with a cut-off argument, one easily extracts also that the limit exists a.s. also under the weaker hypothesis that $\mathbb{E}\left[\log_+ |Y_1|\right] < \infty$, respectively $\mathbb{E}\left[\log_- |Y_1|\right] < \infty$, except that the limit is going to be $-\infty$ unless also the expectation of the negative part is finite (respectively: the limit is going to be $+\infty$ unless also the expectation of the positive part is finite).

The situation with matrix products is more involved, but it is well understood. Nevertheless, the result has a fundamental difference with respect to the case of products of random variables: the limit γ is much less explicit than what we have in (1.2). We will show that the limit can be expressed in terms of the invariant probability of a suitable positive recurrent Markov chain on a continuum space. Except for a few very particular cases (most of them rather *pathological*), one cannot make this invariant probability explicit.

It is however rather easy to compute with very high precision Lyapunov exponents: so we look at a couple of particular cases from the numerical viewpoint before putting the hands on the theory, also to highlight phenomena that are rather surprising.

A first example. Consider for $a > 0$ the matrices

$$M_1 = M_1(a) := a \begin{pmatrix} 1 & 1/5 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad M_2 = M_2(a) := a \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}, \tag{1.3}$$

and, given $(X_j)_{j=1,2,\ldots}$ IID Bernoulli random variables of parameter $1/2$ (fair coin tossing), we define $Y_j := X_j M_1 + (1 - X_j) M_2$.

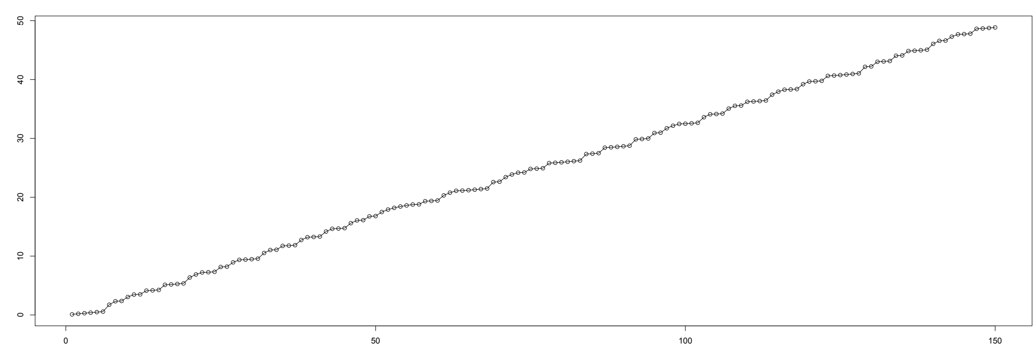

FIGURE 1.1. The case in which we select $M_1(1)$ and $M_2(1)$, given in (1.3), by fair coin tossing (qualitatively the result would be the same also if the coin was biased, as long as not totally biased!): see the definition of (Y_j) right below (1.3). We plot the logarithm of the operator norm of $Y_n Y_{n-1} \ldots Y_1$ as a function of n, up to $n = 150$ (the norm on $\mathbb{R}^2$ is the Euclidean norm). One can of course go to a much larger value of n and estimate that the Lyapunov exponent γ is equal to $0.32\ldots$. In particular, $\exp(\gamma) > 4/3$. We do not discuss here numerical methods and how to control errors.

Of course both $M_1(a)$ and $M_2(a)$ have (both) eigenvalues equal to a: so we start by remarking that, in view of what we remarked numerically in the caption of Figure 1.1, if we choose $a \in [3/4, 1)$ we are are going to have that $\gamma > 0$, i.e. that $\|Y_n Y_{n-1} \ldots Y_1\|$ tends to infinity exponentially fast, even if $\|M_1(a)^n\|$ and $\|M_2(a)^n\|$ tend to zero (exponentially fast). This is (in my view) an instance of what is now often called *Parrondo's paradox* (but the phenomenon in random matrices was known since much earlier): that one can win by switching randomly (or also periodically) between two loosing game strategies.

We have mentioned that $\|Y_n Y_{n-1} \ldots Y_1\|$ and $\|Y_1 Y_2 \ldots Y_n\|$ behave in the same way, at least in the sense of Laplace asymptotic behavior. But $(Y_n Y_{n-1} \ldots Y_1)_n$ and $(Y_1 Y_2 \ldots Y_n)_n$ are very different processes! In fact:

- $(Y_nY_{n-1}\dots Y_1)_n$ is really a random walk on the matrices! Even simpler and more useful: consider
$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} := Y_nY_{n-1}\dots Y_1 \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \tag{1.4}$$
for example for $x_0 > 0$ and $y_0 > 0$. Then also (x_n, y_n) is a point in the first quadrant and (x_0, y_0), $(x_1, y_1), \dots$ is a (time homogeneous) Markov chain on $(0,\infty)^2$. It is actually an elementary exercise to see $(\arctan(y_n/x_n))_n$ is a Markov chain too, this time on $(0, \pi/2)$. We will see that the most important feature of the random walk $(Y_nY_{n-1}\dots Y_1)_n$ is captured by the Markov chain $(\arctan(y_n/x_n))_n$. We plot in Figure 1.2 a numerical approximation of the invariant measure of this Markov chain. So $(\arctan(y_n/x_n))_n$ oscillates wildly! And $(Y_nY_{n-1}\dots Y_1)_n$ is even wilder (because it grows too).
- $(Y_1Y_2\dots Y_n)_n$ is not at all a random walk: we already know that it grows, but in fact it grows *in one specific (random) direction*. Once again, this can be understood much better by applying the matrix to a fixed vector (x_0, y_0) and by considering the arising (x_n, y_n) – note that in the previous item (x_n, y_n) denoted a different process – as we can appreciate by looking at Figure 1.3 where we plot $(\arctan(y_n/x_n))_n$. We insist that this is not a Markov process.

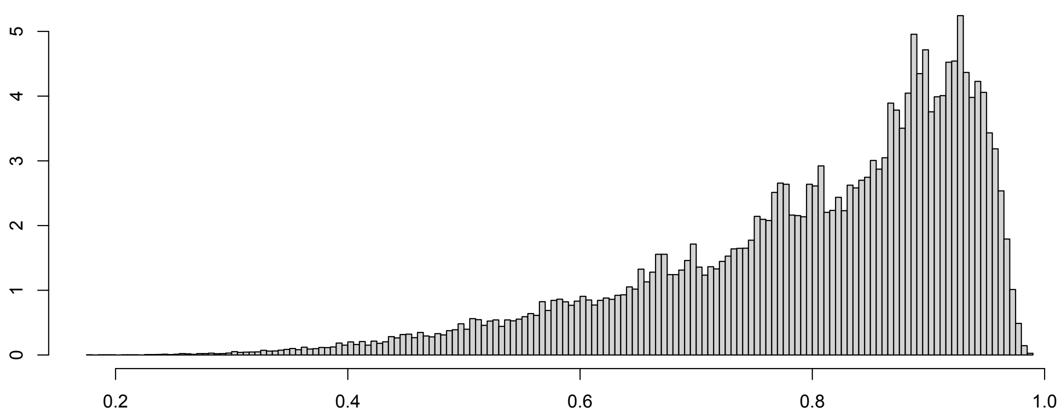

FIGURE 1.2. The invariant measure of the Markov process $(2\arctan(y_n/x_n)/\pi)$ plotted by letting the process run up to $n = 10^7$. The roughness is probably not due to numerical limitations: the measure does not appear to have a smooth density.

REMARK 1.1. *In spite of the fact that the invariant probability may not have a smooth density, see Fig. 1.2, the situation for the choice* (1.3) *is definitely much better than what we have for example with the choice*
$$M_1 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad M_2 := a \begin{pmatrix} 1 & 0 \\ \sqrt{3} & 1 \end{pmatrix}. \tag{1.5}$$
In fact, it is easy to realize that M_1 maps the first quadrant to the sector of the first quadrant with angle from 0 to $\pi/4$, while M_2 maps the first quadrant to the sector

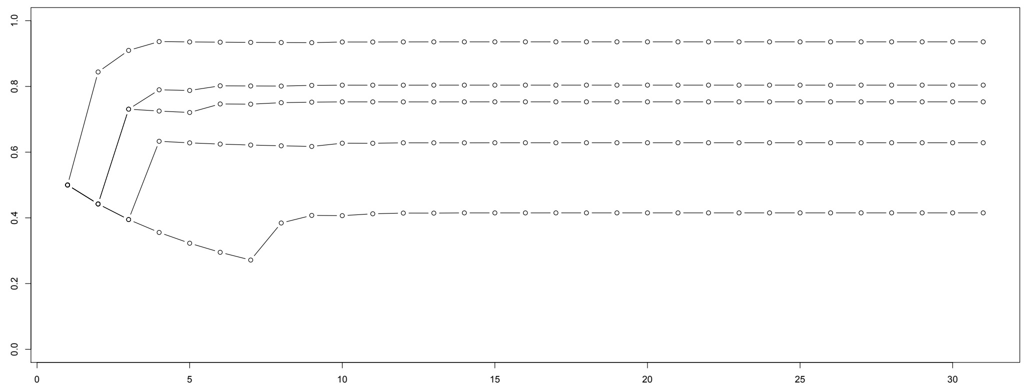

FIGURE 1.3. Plot of five realizations of the process $(2\arctan(y_n/x_n)/\pi)$ in the case in which we are considering $Y_1 Y_2 \ldots Y_n$, so the process is not Markovian. It converges very quickly to a random limit point. We will see that in this particular case the law of this invariant limit point is the invariant measure that appears in Fig. 1.2, but this is just an artefact of the symmetries of this specific example.

with angle from $\pi/3$ to $\pi/2$. So the invariant probability will not be supported on the sector from $\pi/4$ to $\pi/3$. In fact, the support of the invariant probability is not obvious.

A second example. We now consider

$$M_j := \begin{pmatrix} E + \varepsilon U_j & -1 \\ 1 & 0 \end{pmatrix}, \tag{1.6}$$

with (U_j) an IID sequence of $U(-1,1)$ random variables. The eigenvalues of M_j are equal to $(E + \varepsilon U_j \pm \sqrt{(E + \varepsilon U_j)^2 - 4})/2$. They are complex conjugate and of norm one if $-2 < E + \varepsilon U_j < 2$: this is the case of Fig. 1.4 and Fig. 1.5: once again we are in a case in which none of the matrices in the ensemble grows when when multiplied by itself, but the random product has positive Lyapunov exponent (we will prove this in full generality, but it is readily appreciated from simulations)! One can of course consider also the case $|E + \varepsilon U_j| > 2$ in which one of the two eigenvalues is larger than one, and the other are smaller (their product is equal to 1). Also in this case the Lyapunov exponent is positive, but this is less surprising.

1.2. Products of IID random matrices: the top Lyapunov exponent

Let $(Y_k) = (Y_k)_{k \in \mathbb{N}}$ be an IID sequence of random matrices with real entries on the space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that Y_1 is almost surely in $GL_2 := GL_2(\mathbb{R})$, that is $\mathbb{P}(\det(Y_1) \neq 0) = 1$, where $\det(\cdot)$ is the determinant of $\cdot$. We use $\mathcal{M}_2 := \mathcal{M}_2(\mathbb{R})$ for the set of all 2×2 matrices with real entries. Our minimal assumption is that

$$\mathbb{E}\left[\log_+ \|Y_1\|\right] < \infty. \tag{1.7}$$

$\|\cdot\|$ is the operator norm on $\mathcal{M}_2$, i.e. $\|M\| = \sup_{x:\|x\|=1} \|Mx\|$, and depends on the norm $\|\cdot\|$ we choose of $\mathbb{R}^2$. Even if this will be a minor detail because of the

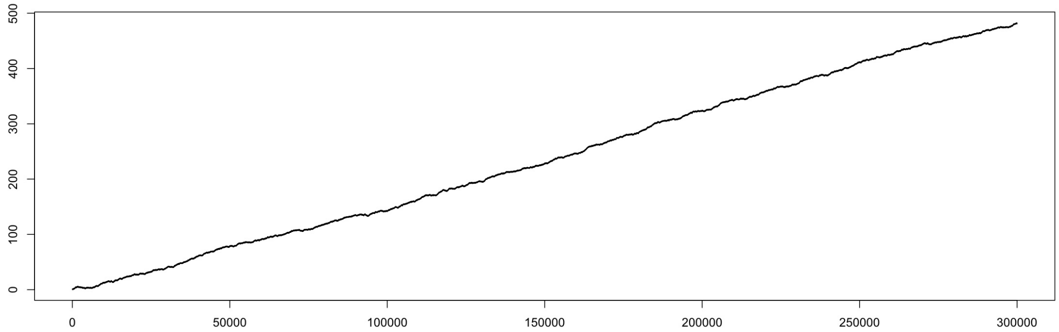

FIGURE 1.4. Plot of the norm of the product in the case of (1.6), with $E = 1.7$ and $\varepsilon = 0.1$.

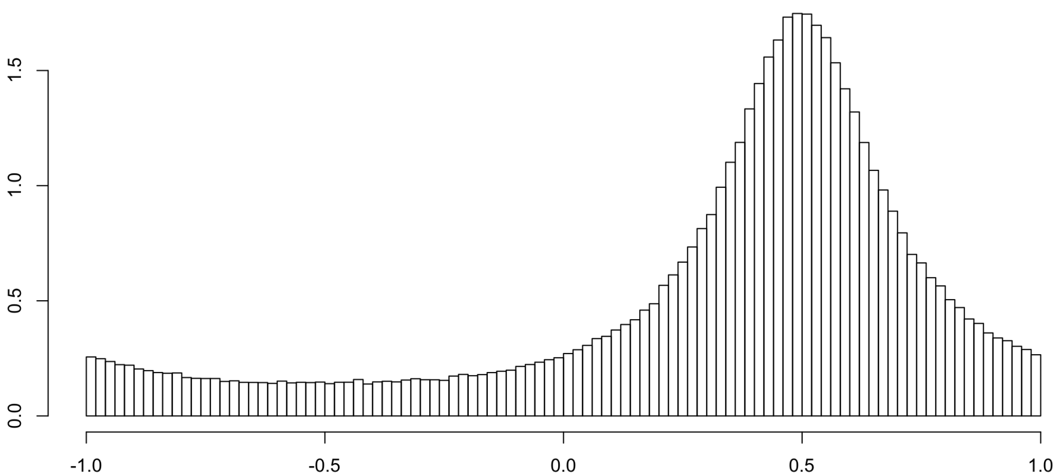

FIGURE 1.5. Invariant probability for the process considered in Fig. 1.4: once again we plot the angle process times $2/\pi$. In this case the matrices do not have non negative entries so the invariant measure is supported also on negative angles. Moreover, the invariant probability appears to have a rather regular density.

equivalence of the norms on $\mathbb{R}^2$, we make and keep the choice $\|x\| := \sqrt{x_1^2 + x_2^2}$. So $\|M\|$ is equal to the square root of the largest eigenvalue of the symmetric matrix M^*M (M^* is the transpose of M). Of course all the eigenvalues of M^*M are non negative: $\langle x, M^*Mx\rangle = \|Mx\|^2 \geq 0$ and we have introduced $\langle x, y\rangle = x_1y_1 + x_2y_2$ for the scalar product.

Since $\|\cdot\|$ is sub-multiplicative, i.e. $\|AB\| \leq \|A\|\|B\|$, and since the law of $(Y_{k+m})_{k\in\mathbb{N}}$ does not depend on $m = 0, 1, \ldots$, we see that, with the notation $P_n := Y_n \ldots Y_1$ for $n \in \mathbb{N}$ and P_0 is the identity matrix, we have

$$\mathbb{E} \log \|P_{n+m}\| \leq \mathbb{E} \log \|Y_{n+m} \ldots Y_{n+1}\| + \mathbb{E} \log \|Y_n \ldots Y_1\| = \mathbb{E} \log \|P_m\| + \mathbb{E} \log \|P_n\| . \tag{1.8}$$

This means that the sequence $(\mathbb{E} \log \|Y_n \ldots Y_1\|)_n$ is sub-additive. We can therefore apply:

Lemma 1.2. *If (a_n) is a sequence of numbers in $\mathbb{R} \cup \{-\infty\}$ such that $a_{n+m} \le a_n + a_m$ for every $n, m \in \mathbb{N} \cup \{0\}$, then the sequence (a_n/n) has a limit and*

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in \mathbb{R} \cup \{-\infty\} . \tag{1.9}$$

We can now introduce:

Definition 1.3. *We call*

$$\gamma := \lim_n \frac{1}{n} \mathbb{E} \log \|Y_n \dots Y_1\| \in \mathbb{R} \cup \{-\infty\} , \tag{1.10}$$

top Lyapunov exponent associated to the sequence (Y_n). The limit coincides with the infimum of the sequence.

We remark that, in our IID context, γ depends only on the law of Y_1. It is straightforward to see that it does not depend on the choice of the norm we have chosen in $\mathbb{R}^2$, nor on the matrix norm (which may not be sub-multiplicative).

Now we aim at an analog of the strong law of large numbers. Namely, we want to show that there is no need to take the expectation in (1.10). This result is an immediate consequence of the probabilistic (or ergodic) version of Lemma 1.2 (i.e., Birkhoff sub-additive Ergodic Theorem), but we provide a self-contained approach with which we will also develop a number of tools that will be useful later on. We will nevertheless rely on the (additive) Birkhoff Ergodic Theorem, i.e. Theorem A.1.

The argument we present is based on the elementary observation that

$$\begin{aligned} \log \|Y_n \dots Y_1\| &= \log \frac{\|Y_n \dots Y_1\|}{\|Y_{n-1} \dots Y_1\|} + \log \|Y_{n-1} \dots Y_1\| \\ &= \sum_{k=1}^{n} \log \frac{\|Y_k \dots Y_1 Y_0\|}{\|Y_{k-1} \dots Y_1 Y_0\|} , \end{aligned} \tag{1.11}$$

where we have introduced Y_0 which is the identity matrix.

This highlights the presence of a general structure: with $\mathbb{G} := GL_2$ and $\mathbb{B} := \{M \in \mathcal{M}_2 : \|M\| = 1\}$ we set for $M \in \mathbb{G}$ and $A \in \mathbb{B}$

$$\sigma(M, A) := \log \|MA\| , \tag{1.12}$$

so σ is an *additive cocycle* in the sense that for $M_1, M_2 \in \mathbb{G}$ and with the notation $M \cdot A := MA/\|MA\|$ we have

$$\sigma(M_2 M_1, A) = \sigma(M_2, M_1 \cdot A) + \sigma(M_1, A) . \tag{1.13}$$

In particular (1.11) can be rewritten as

$$\sigma(Y_n \dots Y_1, Y_0) = \sum_{k=1}^{n} \sigma(Y_k, Y_{k-1} \dots Y_1 \cdot Y_0) , \tag{1.14}$$

with $Y_{k-1} \dots Y_1 \cdot Y_0 = Y_0$ if $k = 1$.

It is useful to introduce some probability notation: if μ is a probability on $\mathbb{G}$ and ν a probability on $\mathbb{B}$, then $\mu \star \nu$ is the law of $\mathrm{M} \cdot \mathrm{A}$ if M is a random variable with law μ and A is a random variable with law ν, and M and A are independent. More explicitly: for every $f \in \mathbb{L}^\infty(\mathbb{B};\mathbb{R})$

$$\int_{\mathbb{B}} f \,\mathrm{d}(\mu \star \nu) \;=\; \int_{\mathbb{G}} \int_{\mathbb{B}} f\left(M \cdot A\right) \mu(\,\mathrm{d}M)\nu(\,\mathrm{d}A)\,. \tag{1.15}$$

Note that in essentially the same way we can define also $\mu_1 \star \mu_2$ with μ_1 and μ_2 two probabilities on $\mathbb{G}$. Moreover these definitions can be generalized to the case in which $\mathbb{G}$ is a topological group that acts continuously on the topological space $\mathbb{B}$. We will always keep the choice $\mathbb{G} = GL_2$, but in the next sections we will work also with a different space $\mathbb{B}$.

We say that ν, probability on $\mathbb{B}$, is $\mu-$invariant, with μ a probability on $\mathbb{G}$, if $\mu \star \nu = \nu$.

Proposition 1.4. $(Y_k)_{k\in\mathbb{N}}$ *is an IID sequence of* $\mathbb{G}$ *valued random variables with* $Y_1 \sim \mu$ *and* ν *is a* μ*-invariant law on* $\mathbb{B}$*. Moreover* $\sigma : \mathbb{G} \times \mathbb{B} \to \mathbb{R}$ *is an additive cocycle such that* $\int_{\mathbb{G}} \int_{\mathbb{B}} |\sigma(M, A)| \mu(\,\mathrm{d}M)\nu(\,\mathrm{d}A) < \infty$*. Then the sequence*

$$\left(\frac{1}{n}\sigma\left(Y_n(\omega)\ldots Y_1(\omega), A\right)\right)_{n\in\mathbb{N}} \tag{1.16}$$

converges $\mathbb{P} \otimes \nu(\,\mathrm{d}(\omega, A))$*-a.s. and in* $\mathbb{L}^1(\mathbb{P} \otimes \nu)$*.*

Proof. This is a matter of showing that this statement is a direct consequence of Birkhoff Ergodic Theorem (Theorem A.1). For this we work on the canonical space $\mathbb{G}^{\mathbb{N}}$ instead of Ω. We set $E := \mathbb{G}^{\mathbb{N}} \times \mathbb{B}$ and we introduce the translation operator $\theta : E \to E$ defined by

$$\theta\left(\left(M_k\right)_{k\in\mathbb{N}}, A\right) \;=\; \left(\left(M_{k+1}\right)_{k\in\mathbb{N}}, M_1 \cdot A\right)\,. \tag{1.17}$$

By using the independence assumptions and the fact that ν is μ-invariant we readily see that θ preserves the probability $\mu^{\otimes\mathbb{N}} \otimes \nu$ on E. Moreover, as pointed out in (1.11)-(1.14), we have

$$\sigma\left(M_n \ldots M_1, A\right) \;=\; \sum_{k=1}^{n} \sigma\left(M_k, M_{k-1}\ldots M_1 \cdot A\right) \;=\; \sum_{k=0}^{n-1} F\left(\theta^k\left((M_j)_{j\in\mathbb{N}}, A\right)\right)\,, \tag{1.18}$$

where in the last step we have introduced the definition $F((M_k), A) := \sigma(M_1, A)$. By the Birkhoff Ergodic Theorem the proof is complete. □

We are now ready to state:

Theorem 1.5 (Furstenberg-Kesten). (Y_k) *are IID* $\mathbb{G}$*-valued random variables and* $\mathbb{E}\log_+ \|Y_1\| < \infty$*. Then* $\lim_n (1/n) \log \|Y_n \ldots Y_1\| = \gamma$ $\mathbb{P}$*-a.s..*

Proof. The upper bound follows easily from the super-multicative property of the matrix norm we use. In fact, choose $p \in \mathbb{N}$ and write $n = mp + r$ with $r \in \{0, \ldots, p-1\}$. Then

$$\frac{1}{n} \log \|Y_n \ldots Y_1\| \le \frac{1}{n} \sum_{k=0}^{m-1} \log \|Y_{p(k+1)} \ldots Y_{pk+1}\| + \frac{1}{n} \log \|Y_{mp+r} \ldots Y_{mp+1}\| . \quad (1.19)$$

The superior limit of the second addendum in the right-hand side is non positive because $\log \|Y_{mp+r} \ldots Y_{mp+1}\| \le \sum_{j=1}^{r} \log_+ \|Y_{mp+j}\|$ and this last term vanishes a.s. as $n \to \infty$ because if (X_j) are IID real random variables with $X_1 \in \mathbb{L}^1$, then $\lim_n X_n/n = 0$ a.s.. We are left with the first term in the right-hand side of (1.19) to which we can apply the Kolmogorov law of Large Numbers for IID random variables and we obtain that a.s. for every $p \in \mathbb{N}$

$$\limsup_n \frac{1}{n} \log \|Y_n \ldots Y_1\| \le \frac{1}{p} \mathbb{E} \left[\log \|Y_p \ldots Y_1\| \right] . \quad (1.20)$$

By taking the infimum over p we obtain that a.s.

$$\limsup_n \frac{1}{n} \log \|Y_n \ldots Y_1\| \le \gamma . \quad (1.21)$$

We are left with the lower bound and for this we can and do assume $\gamma > -\infty$, otherwise there is nothing to prove.

Remark 1.6. *We note that* $\gamma = \inf_n (1/n) \mathbb{E} \log \|Y_n \ldots Y_1\| \le \mathbb{E} \log \|Y_1\|$ *so* $\mathbb{E} \log \|Y_1\| > -\infty$*, that is* $\mathbb{E} |\log \|Y_1\|| < \infty$*. In other terms, we could go ahead in the proof assuming also that* $\mathbb{E} \log_- \|Y_1\| < \infty$*, but this does not seem to be of much help. In any case, this remark contains the interesting information that, under the assumption* $\mathbb{E} \log_+ \|Y_1\| < \infty$*,* $\gamma = -\infty$ *if and only if* $\mathbb{E} \log_- \|Y_1\| = \infty$*.*

Always with the convention that Y_0 is the identity matrix, with $m = \delta_{Y_0}$ (so m is concentrated on the identity matrix) and with $\nu_n := (1/n) \sum_{k=0}^{n-1} \mu^{\star k} \star m$ we write

$$\begin{aligned} \frac{1}{n} \mathbb{E} \log \|Y_n \ldots Y_1\| &= \frac{1}{n} \mathbb{E} \left[\sigma \left(Y_n \ldots Y_1, Y_0 \right) \right] \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E} \left[\sigma \left(Y_{k+1}, Y_k \ldots Y_1 \cdot Y_0 \right) \right] \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \int_{\mathbb{G}} \mathbb{E} \left[\sigma \left(M, Y_k \ldots Y_1 \cdot Y_0 \right) \right] \mu(\mathrm{d}M) \qquad (1.22) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma \left(M, A \right) \mu(\mathrm{d}M) (\mu^{\star k} \star m)(\mathrm{d}A) \\ &= \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma(M, A) \mu(\mathrm{d}M) \nu_n(\mathrm{d}A) . \end{aligned}$$

REMARK 1.7. *Note that the computation in* (1.22) *works also if* Y_0 *is a random variable independent of* $(Y_k)_{k\in\mathbb{N}}$ *and taking values in* $\mathbb{B}$*: of course in the very first term one has to replace* $\|Y_n \dots Y_1\|$ *with* $\|Y_n \dots Y_1 Y_0\|$ *and* m *is the law of* Y_0*. And in the special case in which* m *is* μ *invariant*

$$\frac{1}{n}\mathbb{E}\left[\log \|Y_n \dots Y_1 Y_0\|\right] \,=\, \int_{\mathbb{B}}\int_{\mathbb{G}} \sigma(M,A)\mu(\,\mathrm{d}M)m(\,\mathrm{d}A)\,, \tag{1.23}$$

for every n.

Now we remark that (ν_n) is a sequence of probability mesures on a compact space. It is therefore a relatively compact sequence (for the standard convergence of probability measures) and, thanks to the Césaro mean structure of ν_n

$$\mu \star \nu_n \,=\, \nu_n + \frac{1}{n}\left(\mu^{\star n} \star m - m\right)\,, \tag{1.24}$$

one readily sees that any limit point ν is μ-invariant.

Now we claim that for any limit point ν

$$\sigma \in \mathbb{L}^1(\mu\otimes\nu) \quad\text{and}\quad \int_{\mathbb{B}}\int_{\mathbb{G}} \sigma(M,A)\mu(\,\mathrm{d}M)\nu(\,\mathrm{d}A) \,\ge\, \gamma\,. \tag{1.25}$$

This suffices to complete the proof because by Proposition 1.4 we know that there exists an $\mathbb{L}^1(\mathbb{P}\otimes\nu)$ random variable $(\omega, A)\mapsto \Phi(\omega,A)$ in $\mathbb{L}^1(\mathbb{P}\otimes\nu)$ such that a.s. and in $\mathbb{L}^1$

$$\Phi(\omega, A) \,=\, \lim_n \frac{1}{n}\log\|Y_n \dots Y_1 A\|\,, \tag{1.26}$$

so, in particular, $\int_{\mathbb{B}}\int_{\Omega}\Phi(\omega,A)\mathbb{P}(\,\mathrm{d}\omega)\nu(\,\mathrm{d}A) = \int_{\mathbb{B}}\int_{\mathbb{G}}\sigma(M,A)\mu(\,\mathrm{d}M)\nu(\,\mathrm{d}A)$ because the $\mathbb{P}\otimes\nu$ expectation of what we are taking the limit of in the right-hand side of (1.26) is equal for every n to $\int_{\mathbb{B}}\int_{\mathbb{G}}\sigma(M,A)\mu(\,\mathrm{d}M)\nu(\,\mathrm{d}A)$, as pointed out in Remark 1.7. But, since $\|A\|=1$ ν-a.s., we have

$$\frac{1}{n}\log\|Y_n \dots Y_1 A\| \,\le\, \frac{1}{n}\log\|Y_n \dots Y_1\|\,, \tag{1.27}$$

so, by (1.21), we see the limit of the left-hand side in (1.27) is a.s. bounded by γ. That is, almost surely $\Phi(\omega,A)\le\gamma$. But if we combine this with the second statement in (1.25) we readily obtain that $\Phi(\omega,A)=\gamma$ $\mathbb{P}\otimes\nu$-a.s.. Now it suffices to combine (1.26) and (1.27) to obtain the desired lower bound:

$$\gamma \overset{\mu\otimes\nu\text{-a.s}}{=} \lim_n \frac{1}{n}\log\|Y_n \dots Y_1 A\| \,\le\, \liminf_n \frac{1}{n}\log\|Y_n \dots Y_1\|\,. \tag{1.28}$$

En passant, we observe that this argument implies also that the inequality in the second statement in (1.25) implies the equality.

We are left with the proof of the claim (1.25) and we proceed by considering first σ_+ and then σ_-.

We introduce the non negative map $A\mapsto \int_{\mathbb{G}}\sigma_+(M,A)\mu(\,\mathrm{d}M)$ with domain $\mathbb{B}$. We will now show that this map is continuous (and therefore also bounded, because $\mathbb{B}$ is compact, but an explicit upper bound is given too). In fact for every $M\in\mathbb{G}$ we have that $A\mapsto\|MA\|$ is continuous and $\|MA\|\le\|M\|$. Therefore the non negative function $A \mapsto \sigma_+(M,A) = \log_+ \|MA\|$ is continuous and bounded above by $\log_+ \|M\|$. By $\int_{\mathbb{G}} \log_+ \|M\| \mu(\mathrm{d}M) < \infty$ and the Dominated Convergence we obtain that $A \mapsto \int_{\mathbb{G}} \sigma_+(M,A)\mu(\mathrm{d}M)$ is continuous. Therefore if $\nu_{n_j} \Rightarrow \nu$ we have that

$$\begin{aligned}\lim_j \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma_+(M,A)\mu(\mathrm{d}M)\nu_{n_j}(\mathrm{d}A) \,&=\, \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma_+(M,A)\mu(\mathrm{d}M)\nu(\mathrm{d}A) \\ &\le\, \int_{\mathbb{G}} \log_+ \|M\|\mu(\mathrm{d}M) = \mathbb{E}\left[\log_+ \|Y_1\|\right] < \infty.\end{aligned} \tag{1.29}$$

For the negative part we can apply the same argument if instead of working directly with $\log_-(\cdot)$ we introduce a cut-off $L > 0$ to apply the Dominated Convergence:

$$\lim_j \int_{\mathbb{B}} \int_{\mathbb{G}} \left(\sigma_-(M,A) \wedge L\right) \mu(\mathrm{d}M)\nu_{n_j}(\mathrm{d}A) \,=\, \int_{\mathbb{B}} \int_{\mathbb{G}} \left(\sigma_-(M,A) \wedge L\right) \mu(\mathrm{d}M)\nu(\mathrm{d}A)\,. \tag{1.30}$$

Therefore

$$\liminf_j \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma_-(M,A)\mu(\mathrm{d}M)\nu_{n_j}(\mathrm{d}A) \,\ge\, \int_{\mathbb{B}} \int_{\mathbb{G}} \left(\sigma_-(M,A) \wedge L\right) \mu(\mathrm{d}M)\nu(\mathrm{d}A)\,, \tag{1.31}$$

and the cut-off L can be removed by Monotone Convergence.

By putting together positive and negative parts we therefore reach

$$\limsup_j \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma(M,A)\mu(\mathrm{d}M)\nu_{n_j}(\mathrm{d}A) \,\le\, \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma(M,A)\mu(\mathrm{d}M)\nu(\mathrm{d}A)\,, \tag{1.32}$$

in which the right-hand side is well defined even if a priori it could still be equal to $-\infty$ (and it is bounded above by $\mathbb{E} \log_+ \|Y_1\| =: c$. But by (1.22) and Definition 1.3 the superior limit in the left-hand side is a limit and it is equal to γ, so the second claim in (1.25), the lower bound, is proven. Moreover we have also obtained

$$\int_{\mathbb{B}} \int_{\mathbb{G}} \sigma_+(M,A)\mu(\mathrm{d}M)\nu(\mathrm{d}A) \,\le\, c \quad \text{and} \quad \int_{\mathbb{B}} \int_{\mathbb{G}} \sigma_-(M,A)\mu(\mathrm{d}M)\nu(\mathrm{d}A) \,\le\, c - \gamma\,, \tag{1.33}$$

which completes the proof of (1.25). Therefore Theorem 1.5 is proven too. □

1.3. The Furstenberg formula

In this section we work with a different space $\mathbb{B}$ on which $\mathbb{G} = GL_2(\mathbb{R})$ acts: $\mathbb{B}$ is the projective space $P(\mathbb{R}^2)$, i.e. for $x \in \mathbb{R}^2 \setminus \{0\}$ there is a unique $\overline{x} = x\mathbb{R} \in P(\mathbb{R}^2)$ which we identify ($\cong$) also with $x/\|x\| \cong -x/\|x\|$. $\mathbb{G}$ acts on $\mathbb{B}$ as $M \cdot \overline{x} = \overline{Mx} \cong Mx/\|Mx\|$. Of course given $\overline{x}$ there are plenty of x such that $\overline{x} = x\mathbb{R}$, but in practice this lack of uniqueness does not give problems and, given $\overline{x}$, with some abuse of notation we will denote by x one of these choices. In this section and context we will work with the additive cocycle on $\mathbb{G} \times \mathbb{B}$

$$\sigma(M, \overline{x}) \,:=\, \log \frac{\|Mx\|}{\|x\|}\,. \tag{1.34}$$

$\mathbb{B}$ will be viewed as a metric space with respect to the distance

$$\mathbf{d}\left(\overline{x},\overline{y}\right) := \frac{|\det(x,y)|}{\|x\|\|y\|}, \tag{1.35}$$

where $\det(x,y) = x_1y_2 - x_2y_1$ is the determinant of the matrix with first column equal to x and second column equal y. Moreover $\mathbf{d}(\overline{x},\overline{y})$ coincides with the absolute value of the sine of the angle between the two lines (or rays) $\overline{x}$ and $\overline{y}$.

THEOREM 1.8 (Furstenberg). *We assume that* $(Y_n)_{n\in\mathbb{N}}$ *is an IID sequence of* $\mathbb{G}$*-valued random variables on which we assume that* $\mathbb{P}(|\det(Y_1)| = 1) = 1$ *and that* $\mathbb{E}[\log(\|Y_1\|)] < \infty$. *Moreover we call* μ *the law of* Y_1 *and we assume*

(a) that the smallest closed subgroup G_μ *of* $\mathbb{G}$ *that contains the support of* μ *is not compact;*

(b) that μ *is irreducible in the sense that no finite family of lines (or rays) is stable under the action of all the elements of* G_μ.

Then there exists a unique probability ν *on* $\mathbb{B}$ *which is* μ*-invariant and we have* $\nu(\{\overline{x}\}) = 0$ *for every* $\overline{x} \in \mathbb{B}$. *Moreover*

(1) For every $x \in \mathbb{R}^2 \setminus \{0\}$ *we have that a.s.*

$$\lim_n \frac{1}{n}\log\|Y_n\dots Y_1 x\| = \gamma, \tag{1.36}$$

where γ *is the top Lyapunov exponent of* (Y_n).

*(2) We have (*Furstemberg formula *and positivity of* γ*)*

$$\gamma = \int_{\mathbb{B}}\int_{\mathbb{G}} \log\frac{\|Mx\|}{\|x\|}\mu(\mathrm{d}M)\nu(\mathrm{d}\overline{x}) > 0. \tag{1.37}$$

The property that $\nu(\{\overline{x}\}) = 0$ for every $\overline{x} \in \mathbb{B}$ will be simply stated as "ν is diffuse". Note that $\mathbb{P}(|\det(Y_1)| = 1) = 1$ directly yields $\gamma \geq 0$, because it yields that the determinant of $(Y_n \dots Y_1)^* Y_n \dots Y_1$ is one, hence the largest eigenvalue is not smaller than one. For the same reason (or simpler) we have also that $\|Y_1\| \geq 1$: so the hypothesis is really that $\log(\|Y_1\|)$ is in $\mathbb{L}^1$.

Theorem 1.8 and, in particular, Furstenberg formula for products are given for IID matrices in $\mathbb{G}$ with determinant equal to 1 or -1. This result can easily be adapted to the case in which one does not assume that the determinant is 1 or -1. In fact for $M \in \mathbb{G}$ we can introduce $M^\circ := M/\sqrt{|\det(M)|}$ so

$$\frac{1}{n}\log\|Y_n\dots Y_1\| = \frac{1}{n}\log\|Y_n^\circ\dots Y_1^\circ\| + \frac{1}{2n}\sum_{k=1}^n \log|\det(Y_j)|, \tag{1.38}$$

and if we call $\lambda_+(M)$ and $\lambda_-(M)$ respectively the largest and the smallest eigenvalue of M^*M, then $\det(Y_1) = \sqrt{\det(Y_1^*Y_1)} = \sqrt{\lambda_+(Y_1)\lambda_-(Y_1)}$ and $\|Y_1^\circ\|^2 = \sqrt{\lambda_+(Y_1)/\lambda_-(Y_1)}$. Since we will work under the assumption that $\mathbb{E}\|\log Y_1^\circ\| < \infty$, on the original matrix we are actually requiring that $\mathbb{E}\log(\lambda_+(Y_1)/\lambda_-(Y_1)) < \infty$ and that $\mathbb{E}|\log(\lambda_+(Y_1)\lambda_-(Y_1))| < \infty$.

We now state and prove a series of results that hold under (a subset of) the hypotheses of Theorem 1.8.

1.3.1. Irreducibility implies that μ-invariant probabilities are diffuse. The existence of a μ-invariant probability ν on $\mathbb{B}$ is straightforward by a Césaro mean argument because $\mathbb{B}$ is compact.

PROPOSITION 1.9. *Under the hypotheses of Theorem 1.8 we have that if ν is μ-invariant, then ν is diffuse.*

PROOF. Let us assume by absurd that ν is not diffuse, i.e. that there exists $\overline{x} \in \mathbb{B}$ such that $\nu(\{\overline{x}\}) > 0$. Since ν is a probability, for every $\varepsilon > 0$ there is a most a finite number of $\overline{x}$ such that $\nu(\{\overline{x}\}) \geq \varepsilon$ and therefore there exists $\delta > 0$ such that

- $\nu(\{\overline{x}\}) \leq \delta$ for every $\overline{x} \in \mathbb{B}$;
- if we set $S := \{\overline{x} \in \mathbb{B} : \nu(\{\overline{x}\}) = \delta\}$, then $1 \leq |S| < \infty$.

We are going to show that S is stable under the action of G_μ, which is in contrast Hypothesis (b) of Theorem 1.8, so ν is diffuse.

For this we first choose $\overline{y} \in S$ and remark that

$$\delta = \nu(\{\overline{y}\}) = (\mu \star \nu)(\{\overline{y}\}) = \int_{\mathbb{G}} \int_{\mathbb{B}} \mathbf{1}_{\{M \cdot \overline{x} = \overline{y}\}} \nu(\mathrm{d}\overline{x}) \mu(\mathrm{d}M)$$
$$= \int_{\mathbb{G}} \nu\left(\left\{M^{-1} \cdot \overline{y}\right\}\right) \mu(\mathrm{d}M) \leq \delta, \quad (1.39)$$

because one directly verifies that $M \cdot \overline{x} = \overline{y}$ if and only if $\overline{x} = M^{-1} \cdot \overline{y}$ and because $\nu(\{M^{-1} \cdot \overline{y}\}) \leq \delta$ by hypothesis. But this implies that there exists a measurable subset F of $\mathbb{G}$, with $\mu(F) = 1$, such that $\nu(\{M^{-1} \cdot \overline{y}\}) = \delta$ for every $\overline{y} \in S$ and for every $M \in F$. In other words, $M^{-1} \cdot \overline{y} \in S$ for every $\overline{y} \in S$ and for every $M \in F$. Of course we cannot have $M^{-1} \cdot \overline{y} = M^{-1} \cdot \overline{x}$ for $\overline{x} \neq \overline{y}$, that is $\overline{x} \mapsto M^{-1} \cdot \overline{x}$ is injective and, since S is finite, it is also surjective. Hence the inverse function, $\overline{x} \mapsto M \cdot \overline{x}$, is also a bijection from S to itself.

Now we remark that, by the definition of Supp(μ) (the support of μ), i.e. the set of $M \in \mathbb{G}$ for which every neighborhood has positive μ measure, we have that F is dense in Supp(μ). Therefore $M \cdot \overline{y} \in S$ as well as $M^{-1} \cdot \overline{y} \in S$ for every $\overline{y} \in S$ and for every $M \in$ Supp(μ) and, in turn, also for every $M \in G_\mu$. So S is stable under the action of G_μ and the proof is complete.

□

1.3.2. About the action of the transposed matrices. Considering the transposed random variables $X_n := Y_n^*$ does not make a big difference. If (Y_n) satisfies the hypotheses in Theorem 1.8, the same hypotheses are satisfied by (X_n). We state this in a lemma in which we use μ^* for the law of X_1.

LEMMA 1.10. *If (Y_n) satisfies the hypotheses in Theorem 1.8, then the same holds for (X_n). In particular (X_n) is an IID sequence of $\mathbb{G}$ valued random* *variables with absolute value of the determinant equal to one, G_{μ^*} is not compact and there exists no finite union of rays that is stable under the action of G_{μ^*}.*

In particular, Theorem 1.9 applies and any μ^*-invariant probability ν^* is diffuse.

PROOF. The basic facts (value of the determinant and integrability properties) are evident. The non compactness property follows because taking the transpose is a homeomorphism (i.e., bijective continuous map with continuous inverse) from $\mathbb{G}$ to $\mathbb{G}$. So if we take the transpose of the elements in the smallest closed subgroup that contains the support of μ, we find the smallest closed subgroup that contains the support of μ^*, i.e. $G_{\mu^*} = (G_\mu)^*$. Moreover the norm of $M \in \mathbb{G}$ coincides with the norm of M^* because the spectrum of M^*M coincides with the spectrum of MM^*. Since the non compactness property of a closed subset G of $\mathbb{G}$ is just the fact that the $\sup_{M\in G} \|M\| = \infty$ we are done.

The X matrices inherit the irreducibility property because if there exists a finite family of rays that is stable under the action of G_{μ^*}, there exists such a family also for the Y matrices. In fact this statement is symmetric: there exists a finite family of rays that is stable under the action of G_{μ^*} if and only if there exists a finite family of rays that is stable under the action of G_μ. This can be seen by observing that if $M \cdot \overline{x} = \overline{x}$, then $Mx = \lambda x$ for some $\lambda \neq 0$, so $0 = \langle x^\perp, Mx\rangle = \langle M^* x^\perp, x\rangle$, where of course $x^\perp$ is a non zero vector perpendicular to x. Hence $M^* x^\perp = \lambda' x^\perp$ and we have a fixed ray also for the transposed matrix. This takes care of the case in which the family of rays is just given by one ray. In the general case in which there are k rays $\{\overline{x}_1, \ldots, \overline{x}_k\}$ that are fixed for every $M \in G_\mu$ (in reality one can show that k is either one or two) then, since $\overline{x} \mapsto M \cdot \overline{x}$ is injective, for every ray $\overline{x}$ in this family there exists $j \leq k$ such that $M^j \cdot \overline{x} = \overline{x}$. Then, by the previous argument, $(M^*)^j$ fixes the ray $\overline{x}^\perp$. This directly yields that $\{\overline{x}^\perp, M^* \cdot \overline{x}^\perp, \ldots, (M^*)^{j-1} \cdot \overline{x}^\perp\}$ is invariant under the action of M^* and this directly yields that $M^* \cdot \{\overline{x}_1, \ldots, \overline{x}_k\} = \{\overline{x}_1, \ldots, \overline{x}_k\}$ for every $M \in G_\mu$. To wrap it up: if if there exists a finite family of rays that is stable under the action of G_{μ^*}, then the family of the orthogonal rays are stable under the action of G_μ, but such a family does not exist by hypothesis. □

The novelty in working with the transposed matrices is the inversion of the order: $(Y_n \ldots Y_1)^* = X_1 \ldots X_n$. If this is irrelevant from certain viewpoints, notably $\|(Y_n \ldots Y_1)^*\| = \|Y_n \ldots Y_1\|$ so $\gamma = \lim_n (1/n)\mathbb{E} \log \|X_1 \ldots X_n\|$, it is not irrelevant at all from other aspects. In particular, $X_1 \ldots X_n \cdot \overline{x}$ converges in probability (and even a.s. under slightly stronger assumptions): in Proposition 1.11 one can find a result that goes toward this direction. Note that this is definitely false for $Y_n \ldots Y_1 \cdot \overline{x}$: the action of Y_{n+1} will substantially modify the direction of the vector and one can only hope for convergence in law for $Y_n \ldots Y_1 \cdot \overline{x}$.

Here is the main result of Section 1.3.2:

PROPOSITION 1.11. *We assume the validity of the hypotheses in Theorem 1.8. There exists a unique probability ν^* that is μ^*-invariant and a $\mathbb{B}$ valued random* *variable* $\overline{Z}$ *(of law* ν^**) such that* $\mathbb{P}(\mathrm{d}\omega)$*-a.s.*

$$X_1(\omega)\dots X_n(\omega)\nu^* \stackrel{n\to\infty}{\Longrightarrow} \delta_{\overline{Z}(\omega)}\,. \tag{1.40}$$

REMARK 1.12. *Note that, since by Lemma 1.10 the sequences* (Y_n) *and* (X_n) *satisfy the same assumptions, Proposition 1.11 implies also that there exists a unique* μ*-invariant probability* ν *(and, by Proposition 1.9,* ν*, as well as* ν^**, is diffuse.*

A main step in proving Proposition 1.11 is the following statement, that will be proven by a martingale argument.

LEMMA 1.13. *Under the hypotheses in Theorem 1.8 we have that if* ν^* *is a* μ^**-invariant probability then there exists* $\omega \mapsto \nu_\omega$*, measurable map for* Ω *to the probabilities over* $\mathbb{B}$ *(equipped with the weak convergence metric and the corresponding Borel* σ*-algebra) such that* $\mathbb{P}(\mathrm{d}\omega)$*-a.s.*

$$X_1(\omega)\dots X_n(\omega)\nu^* \stackrel{n\to\infty}{\Longrightarrow} \nu_\omega\,, \tag{1.41}$$

and $(\mathbb{P}\otimes\mu^*)(\mathrm{d}(\omega, M))$ *a.s.*

$$X_1(\omega)\dots X_n(\omega)M\nu^* \stackrel{n\to\infty}{\Longrightarrow} \nu_\omega\,. \tag{1.42}$$

Moreover $\int_\Omega\int_{\mathbb{B}} f\,\mathrm{d}\nu_\omega\mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{B}} f\,\mathrm{d}\nu^*$ *for every* $f \in L^\infty(\mathbb{B};\mathbb{R})$*.*

PROOF OF LEMMA 1.13. For $f \in C^0(\mathbb{B};\mathbb{R})$, hence $f \in C^0_b(\mathbb{B};\mathbb{R})$, we introduce the function $F : \mathbb{G} \to \mathbb{R}$ defined by

$$F(M) \,=\, F_f(M) \,:=\, \int_{\mathbb{B}} f\,(M\cdot\overline{x})\,\nu^*(\mathrm{d}\overline{x})\,. \tag{1.43}$$

In analogy with the definition of Y_0, X_0 is the identity matrix. We now define $M_n := X_0X_1\dots X_n$ and claim that $(F(M_n))_{n=0,1,\dots}$ is a martingale with respect to the natural filtration $(\mathcal{F}_n)$ of $(X_n)_{n=0,1,\dots}$. $F(M_n)$ is in fact measurable with respect to $\mathcal{F}_n$, it is a real random variable in $\mathbb{L}^1$ (in fact, in $\mathbb{L}^\infty$) and

$$\begin{aligned}
\mathbb{E}\left[F\left(M_{n+1}\right)\middle|\mathcal{F}_n\right](\omega) \,&=\, \int_{\mathbb{G}} F\left(M_n(\omega)M\right)\mu^*(\mathrm{d}M)\\
&=\, \int_{\mathbb{G}}\int_{\mathbb{B}} f\left(M_n(\omega)M\cdot\overline{x}\right)\mu^*(\mathrm{d}M)\nu^*(\mathrm{d}\overline{x})\\
&=\, \int_{\mathbb{B}} f\left(M_n(\omega)\cdot\overline{y}\right)(\mu^*\star\nu^*)(\mathrm{d}\overline{y})\\
&=\, \int_{\mathbb{B}} f\left(M_n(\omega)\cdot\overline{y}\right)\nu^*(\mathrm{d}\overline{y}) \,=\, F\left(M_n(\omega)\right)\,.
\end{aligned} \tag{1.44}$$

Since $\|F\left(M_n\right)\|_\infty \le \|f\|_\infty$, the Martingale Convergence Theorem guarantees that $(F(M_n))$ converges a.s. and, by the Dominated Convergence Theorem, also in L^1 (in fact, in any L^p, $p\ge 1$), to a random variable that we denote by Q_f. We now use that we can find a sequence (f_k) of elements in $C^0(\mathbb{B};\mathbb{R})$ which is dense in $C^0(\mathbb{B};\mathbb{R})$, so $\mathbb{P}(\mathrm{d}\omega)$ a.s. we have that for every k the sequence $(F_{f_k}(M_n(\omega)))$ converges to $Q_{f_k}(\omega)$

for every ω in a probability one subset Ω_0 of Ω. Hence $(F_f(M_n(\omega)))$ converges to $Q_f(\omega)$ for every $f \in C^0(\mathbb{B};\mathbb{R})$ for every $\omega \in \Omega_0$. Since $\mathbb{B}$ is compact, this means that $(M_n(\omega)\nu^*)$ converges to a limit probability ν_ω for the same ω's, and we have $Q_f(\omega) = \int_{\mathbb{B}} f \,\mathrm{d}\nu_\omega$. By defining ν_ω to be an arbitrary fixed probability measure on $\mathbb{B}$ when $\omega \notin \Omega_0$, one readily verifies the measurability properties of $\omega \mapsto \nu_\omega$. Therefore (1.41) is established.

Note that the martingale property implies that $\int_{\mathbb{B}} f \,\mathrm{d}\nu^* = F_f(M_0) = \mathbb{E}[F_f(M_n)]$ for every n and we have seen that $\lim_n \mathbb{E}[F_f(M_n)] = \mathbb{E}[Q_f] = \int_\Omega \int_{\mathbb{B}} f \,\mathrm{d}\nu_\omega \mathbb{P}(\mathrm{d}\omega)$. Therefore $\int_\Omega \int_{\mathbb{B}} f \,\mathrm{d}\nu_\omega \mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{B}} f \,\mathrm{d}\nu^*$ and the last claim in Lemma 1.13 is established.

We are left with establishing (1.42): we aim at showing that that there exists a measurable subset Ω_f of Ω with $\mathbb{P}(\Omega_f) = 1$ and a measurable subset G_f of $\mathbb{G}$ with $\mu^*(G_f) = 1$ such that

$$\lim_n F_f(M_n(\omega)M) \,=\, Q_f(\omega)\,, \tag{1.45}$$

for every $\omega \in \Omega_f$ and every $M \in G_f$. This suffices because we can find a sequence (f_k) of elements in $C^0(\mathbb{B};\mathbb{R})$ which is dense in $C^0(\mathbb{B};\mathbb{R})$ and of course $\mathbb{P}(\cap_k \Omega_{f_k}) = 1$ as well as $\mu^*(\cap_k G_{f_k}) = 1$

For this we exploit that $(F(M_n))$ is a martingale in $\mathbb{L}^2$ so

$$\sum_{k=1}^{n} \mathbb{E}\left[(F(M_k) - F(M_{k-1}))^2\right] \,=\, \mathbb{E}\left[F(M_n)^2\right] - \mathbb{E}\left[F(M_0)^2\right] \,\le\, \|f\|_\infty^2\,. \tag{1.46}$$

Therefore

$$\sum_{k=1}^{\infty} \mathbb{E}\left[(F(M_k) - F(M_{k-1}))^2\right] \,\le\, \|f\|_\infty^2 \,<\infty\,. \tag{1.47}$$

But since (X_k) is an IID sequence

$$\mathbb{E}\left[(F(M_{k+1}) - F(M_k))^2\right] \,=\, \int_{\mathbb{G}} \mathbb{E}\left[(F(M_kM) - F(M_k))^2\right] \mu^*(\mathrm{d}M)\,, \tag{1.48}$$

hence, by (1.47) and the Fubini-Tonelli Theorem, we have

$$\int_{\mathbb{G}} \mathbb{E}\left[\sum_{k=1}^{\infty} (F(M_kM) - F(M_k))^2\right] \mu^*(\mathrm{d}M) \,<\, \infty\,. \tag{1.49}$$

In particular $\lim_k F(M_k(\omega)M) - F(M_k(\omega)) = 0$ $\mathbb{P}(\mathrm{d}\omega)$ and $\mu^*(\mathrm{d}M)$ almost surely, and this yields (1.45) since $\lim_k F(M_k(\omega)) = Q_f(\omega)$ $\mathbb{P}(\mathrm{d}\omega)$-a.s.. □

In what follows we are going to need to work also with matrices which do not have an inverse. Notably, we need to define Am for some probabilities m on $\mathbb{B}$ and this could be ill defined because defining Am involves computing $A \cdot \overline{x}$ which is not defined if $Ax = 0$. However this is not really a problem if we exclude the trivial case $A \equiv 0$ (i.e., at least one entry of A is non zero) and if we assume that m is diffuse. This is the content of the next lemma.

LEMMA 1.14. *If $A \not\equiv 0$ and m is a diffuse probability on $\mathbb{B} = P(\mathbb{R}^2)$ then Am is the probability defined by*

$$\int_{\mathbb{B}} f\,\mathrm{d}(Am) \,=\, \int_{\{\overline{x}\in\mathbb{B}:\, Ax\neq 0\}} f\left(A\cdot\overline{x}\right) m(\,\mathrm{d}\overline{x})\,, \tag{1.50}$$

for every $f:\mathbb{B}\to\mathbb{R}$ bounded and measurable. Moreover if (A_n) converges to A, then $A_n m \Rightarrow Am$.

PROOF. The key point is to observe that, since $A \not\equiv 0$, there exists at most one ray $\overline{y}$ such that $Ay = 0$. So, $A\cdot\overline{x}$ is ill defined only for $\overline{x} = \overline{y}$. But $m(\{\overline{y}\}) = 0$ because m is diffuse so we can proceed to define the Am as we did. The fact that Am is a probability follows from the properties of the integral in the right-hand side of (1.50): in particular $\int_{\mathbb{B}} \mathrm{d}(Am) = 1$ because $m(\{\overline{y}\}) = 0$.

The second statement holds because if we call $\overline{y}_n$, respectively $\overline{y}$, the ray such that $A_n y_n = 0$, respectively $Ay = 0$ (they may or may not exist), we can perform the limiting (Dominated Convergence) procedure by integrating over $\{\overline{x}\in\mathbb{B}:\, x\neq y$ and $x \neq y_n$ for every $n\}$, since this is a set of m-probability one. □

We are now ready to prove Proposition 1.11.

PROOF OF PROPOSITION 1.11. We exploit the content of Lemma 1.13 and also the notation $M_n = X_1 \ldots X_n$ used in its proof. For ν^* a μ^*-invariant probability – we recall that by Lemma 1.9 ν^* is diffuse – Lemma 1.13 tells us that (for $\omega\in\Omega_0$ and $M\in G_0$, $\mathbb{P}(\Omega_0) = 1$ and $\mu^*(G_0) = 1$) $M_n(\omega)\nu^* \Rightarrow \nu_\omega$ and $M_n(\omega)M\nu^* \Rightarrow \nu_\omega$. Consider then the sequence of norm one matrices $(M_n(\omega)/\|M_n(\omega)\|)$ for $\omega\in\Omega_0$: this is a relatively compact sequence and we call $A(\omega)$ a limit point. Since $\|A(\omega)\| = 1$, then $A(\omega)\not\equiv 0$. Moreover $M_n(\omega)\nu^* = (M_n(\omega)/\|M_n(\omega)\|)\nu^*$. So by Lemma 1.14 we have that for $\omega\in\Omega_0$ and $M\in G_0$

$$A(\omega)\nu^* \,=\, \nu_\omega \;\text{ and }\; A(\omega)M\nu^* \,=\, \nu_\omega\,. \tag{1.51}$$

Let us show that we cannot have $\det(A(\omega))\neq 0$. In fact in this case we would have $M\nu^* = (A(\omega))^{-1}\nu_\omega$ as well as $\nu^* = (A(\omega))^{-1}\nu_\omega$. So $M\nu^* = \nu^*$ for every $M\in G_0$ and this extends to every $M\in$Supp(μ) because G_0 is dense in Supp(μ). But we have also $M\nu^* = \nu^*$ is equivalent to $\nu^* = M^{-1}\nu^*$, so $M\nu^* = \nu^*$ for every $M\in G_\mu$. Therefore we see that

$$G_\mu \,\subset\, \{M\in\mathbb{G}:\, |\det(M)| = 1 \text{ and } M\nu^* = \nu^*\} \,=:\, G_1\,. \tag{1.52}$$

but G_1 is a compact subset of $\mathbb{G}$ and this is impossible because G_μ is not. In fact, if G_1 is not compact we can find $M_n\in G_1$ with $\|M_n\|\to\infty$. Using the relative compactness of $(M_n/\|M_n\|)$ we can extract a subsequence along which this sequence converges to a limit matrix that we call C for which $C\nu^* = \nu^*$ still holds and we have also that $\lim_n \det(M_n/\|M_n\|)\to\det(C)$. But $\det(M_n/\|M_n\|) = \pm 1/\|M_n\|^2\to 0$ so $\det(C) = 0$. But then $C\nu^* = \nu^*$ implies that ν^* is concentrated on one ray and this is impossible because ν^* is diffuse.

We are therefore dealing with $A(\omega)$ of rank one, so there exists one (and only one) ray $\overline{y}(\omega)$ for which $A(\omega)\cdot\overline{y}(\omega)$ is not defined. And there exists also $\overline{Z}(\omega)\in\mathbb{B}$ such that for $\overline{x}\neq\overline{y}(\omega)$ we have $A(\omega)\cdot\overline{x}=\overline{Z}(\omega)$. Therefore for every $\omega\in\Omega_0$

$$\nu_\omega \,=\, A(\omega)\nu^* \,=\, \delta_{\overline{Z}(\omega)} \quad\text{and}\quad X_1(\omega)\ldots X_n(\omega)\nu^* \Longrightarrow \delta_{\overline{Z}(\omega)}\,. \tag{1.53}$$

Lemma 1.13 is telling us also that for f bounded measurable

$$\mathbb{E}\left[f\left(\overline{Z}\right)\right] \,=\, \int_\Omega\int_{\mathbb{B}} f\,\mathrm{d}\nu_\omega \mathbb{P}(\,\mathrm{d}\omega) \,=\, \int_{\mathbb{B}} f\,\mathrm{d}\nu^*\,, \tag{1.54}$$

which means that the law of the $\mathbb{B}$-valued random variable $\overline{Z}$ is ν^*. Remark now the important fact that $A(\omega)$ does not dependent on the choice of ν^*, so $\overline{Z}(\omega)$ does not depend on this choice either: therefore there is a unique μ^*-invariant probability. □

We complete this subsection by giving a deterministic result that helps substantially in going from results on the transposed matrix to the matrix itself.

Lemma 1.15. *Let (A_n) be a sequence of 2×2 matrices such that $|\det(A_n)|=1$ for every n and such that there exists a diffuse probability m on $\mathbb{B}$ and a ray $\overline{z}\in\mathbb{B}$ such that $A_n^*m\Rightarrow\delta_{\overline{z}}$. Then*

(1) $\lim_n\|A_n\|=\infty$*;*

(2) for every $x\in\mathbb{R}^2$

$$\lim_n \frac{\|A_nx\|}{\|A_n\|} \,=\, |\langle x,\overline{z}\rangle|\,, \tag{1.55}$$

where we recall that $\langle\cdot,\cdot\rangle$ is the scalar product in $\mathbb{R}^2$ and in the right-hand side $\overline{z}$ should be read as $z/\|z\|$.

Proof. By passing to a subsequence we can assume $A_n/\|A_n\|\to A$. Since $\|A_n^*\|=\|A_n\|$ and $(A_n^*/\|A_n^*\|)m=A_n^*m$ we have $A^*m=\delta_{\overline{z}}$. If $\det(A)\neq0$ then $m=(A^*)^{-1}\delta_{\overline{z}}=\delta_{(A^*)^{-1}\cdot\overline{z}}$ which is impossible because m is diffuse.

So $\det(A)=0$. But $\pm1/\|A_n\|^2=\det(A_n/\|A_n\|)\to\det(A)=0$ so $\|A_n\|\to\infty$. This proves (1).

For (2) we use that A is of rank one, since it is not identically zero. Therefore $A^*m=\delta_{\mathrm{Im}(A^*)}$ where of course $\mathrm{Im}(A^*)$ is the image of A^* acting on $\mathbb{B}$, but it is also the Image as linear operator on $\mathbb{R}^2$ (a line that goes through the origin). In particular we have $\mathrm{Im}(A^*)=\{\overline{z}\}$ and the kernel of A (as linear operator on $\mathbb{R}^2$) coincides with the orthogonal of $\mathrm{Im}(A^*)$. Therefore if we choose z_1 to be a unit vector in the line $\overline{z}$ and z_2 a unit vector orthogonal to z_1, for $x\in\mathbb{R}^2$

$$Ax \,=\, \langle x,z_1\rangle Az_1+\langle x,z_2\rangle Az_2 \,=\, \langle x,z_1\rangle Az_1\,, \tag{1.56}$$

and

$$1 \,=\, \|A\| \,=\, \max_{x:\|x\|=1}\|Ax\| \,=\, \max_{x:\|x\|=1}|\langle x,z_1\rangle|\,\|Az_1\| \,=\, \|Az_1\|\,. \tag{1.57}$$

Now it suffices to observe that

$$\lim_n \frac{\|A_n x\|}{\|A_n\|} \,=\, \|Ax\| \,=\, |\langle x, z_1\rangle| \, \|Az_1\| \,=\, |\langle x, z_1\rangle| \,. \tag{1.58}$$

□

1.3.3. On the $\mathbb{L}^1$ character of the cocycles and on the positivity of γ. In this section we develop the tools to show that the Lyapunov exponent is non zero. This will go through showing that if the positive part of the cocycle is in $\mathbb{L}^1$ and if the cocycle evaluated on the product of random matrices $Y_n \dots Y_1$ tends to ∞ in a suitable (weak) sense, then the cocycle is in $\mathbb{L}^1$ and the growth is actually proportional to n. We give this result in a rather general framework.

LEMMA 1.16. *(Y_n) is an IID sequence of $\mathbb{G}$-valued random variables with common law μ. We consider an additive cocycle σ on $\mathbb{G} \times \mathbb{B}$. We assume that ν is a μ-invariant probability on $\mathbb{B}$ and that*

$$\int_{\mathbb{B}} \int_{\mathbb{G}} \sigma_+(M, \overline{x}) \mu(\,\mathrm{d}M) \nu(\,\mathrm{d}\overline{x}) \,<\, \infty, \tag{1.59}$$

and that $(\mathbb{P} \otimes \nu)(\,\mathrm{d}(\omega, \overline{x}))$-a.s.

$$\lim_n \sigma\left(Y_n(\omega) \dots Y_1(\omega), \overline{x}\right) \,=\, \infty \,. \tag{1.60}$$

Then $\sigma \in \mathbb{L}^1(\mathbb{P} \otimes \nu)$ and

$$\int_{\mathbb{B}} \int_{\mathbb{G}} \sigma(M, \overline{x}) \mu(\,\mathrm{d}M) \nu(\,\mathrm{d}\overline{x}) \,>\, 0 \,. \tag{1.61}$$

PROOF. Like in the proof of Proposition 1.4, we work on $E = \mathbb{G}^{\mathbb{N}} \times \mathbb{B}$ equipped with the product probability $\Lambda := \mu^{\otimes \mathbb{N}} \otimes \nu$ and θ is the translation operator defined like before, i.e. $\theta((M_k)_{k \in \mathbb{N}}, \overline{x}) = ((M_{k+1})_{k \in \mathbb{N}}, M_1 \cdot \overline{x})$. Note that θ is Λ-preserving.

We focus on $F((M_k)_{k \in \mathbb{N}}, \overline{x}) := \sigma(M_1, \overline{x})$ and (again, like in the proof of Proposition 1.4) we have

$$\sigma\left(Y_n \dots Y_1, \overline{x}\right) \,=\, \sum_{k=1}^{n} \sigma\left(Y_k, Y_{k-1} \dots Y_1 \cdot \overline{x}\right) \,=\, \sum_{k=0}^{n-1} F \circ \theta^k \left((Y_n), \overline{x}\right) \,. \tag{1.62}$$

From now on, for conciseness, we call x an element of E, $\mathcal{F}$ the σ-algebra on E and what we want to show is that if $f : E \to \mathbb{R}$ is a measurable function such that $\int_E f_+ \,\mathrm{d}\Lambda < \infty$ and $S_n(x) := \sum_{k=0}^{n-1} f \circ \theta^k(x) \to \infty$ $\Lambda(\,\mathrm{d}x)$-a.s. (it is practical to set for later on $S_0(x) := 0$), then $f \in \mathbb{L}^1$ and $\int_E f \,\mathrm{d}\Lambda > 0$.

For this we apply the Birkhoff Ergodic Theorem (Corollary A.2) to obtain that $\Lambda(\,\mathrm{d}x)$-a.s.

$$\lim_n \frac{1}{n} S_n(x) \,=\, \mathbf{E}\left[f \,\middle|\, \mathcal{I}\right](x) \,, \tag{1.63}$$

where $\mathbf{E}$ is the expectation with respect to the probability Λ and $\mathcal{I}$ is the σ-algebra of the events $A \in \mathcal{F}$ such that $\Lambda(A \Delta \theta^{-1}(A)) = 0$. Since $S_n(x) \to \infty$ we directly obtain that a.s. $\mathbf{E}[f|\,\mathcal{I}](x) \geq 0$ and therefore $\mathbf{E}[f] = \int_E f \,\mathrm{d}\Lambda \geq 0$. In particular $\int_E f_- \,\mathrm{d}\Lambda \leq \int_E f_+ \,\mathrm{d}\Lambda < \infty$, i.e. $f \in \mathbb{L}^1$.

We are therefore left with showing that $\int_E f \,\mathrm{d}\Lambda > 0$. Let us therefore assume that $\int_E f \,\mathrm{d}\Lambda = 0$ and look for a contradiction. So we start by observing that $0 = \int_E f \,\mathrm{d}\Lambda = \mathbf{E}[\mathbf{E}[f|\,\mathcal{I}]] = 0$ and we have seen that $\mathbf{E}[f|\,\mathcal{I}] \geq 0$ a.s.. So $\mathbf{E}[f|\,\mathcal{I}] = 0$ a.s. and this means that $S_n(x) = o(n)$ a.s..

To show that this is impossible we use (S_n) to build a sequence of random subsets of the real line equipped with the Lebesgue measure λ. This sequence of random sets depends on a parameter $\varepsilon > 0$ and the n^{th} set is the union of the intervals $S_k(x) + [-\varepsilon, \varepsilon] =: I_\varepsilon(S_k(x))$ for k going from 0 to $n-1$. We call $m_n^\varepsilon(x)$ the Lebesgue measure of this set:

$$m_n^\varepsilon(x) \,=\, \lambda\left(\bigcup_{k=0}^{n-1} I_\varepsilon\left(S_k(x)\right)\right) . \tag{1.64}$$

The straightforward rough upper bound

$$m_n^\varepsilon(x) \,\leq\, 2\left(\max_{k=0,\ldots,n-1} |S_k(x)| + \varepsilon\right) \overset{\Lambda(\,\mathrm{d}x)\text{-a.s}}{=} o(n)\,, \tag{1.65}$$

and the even more immediate remark that $m_n^\varepsilon(x) \leq 2\varepsilon n$ lead (via Dominated Convergence) to $\lim_n \mathbf{E}[m_n^\varepsilon]/n = 0$.

Now we play on the structure of S_n: in particular we have $S_k \circ \theta = S_{k+1} - S_1$. Using $\cup_0^n A_k \setminus \cup_1^n A_k = A_0 \setminus \cup_1^n A_k$ and $S_0 = 0$ we obtain

$$\begin{aligned} m_{n+1}^\varepsilon(x) - m_n^\varepsilon(\theta x) \,&=\, \lambda\left(\bigcup_{k=0}^{n} I_\varepsilon\left(S_k(x)\right)\right) - \lambda\left(\bigcup_{k=0}^{n-1} I_\varepsilon\left(S_{k+1}(x) - S_1(x)\right)\right) \\ &=\, \lambda\left(\bigcup_{k=0}^{n} I_\varepsilon\left(S_k(x)\right)\right) - \lambda\left(\bigcup_{k=0}^{n-1} I_\varepsilon\left(S_{k+1}(x)\right)\right) \\ &=\, \lambda\left(\bigcup_{k=0}^{n} I_\varepsilon\left(S_k(x)\right)\right) - \lambda\left(\bigcup_{k=1}^{n} I_\varepsilon\left(S_k(x)\right)\right) \\ &=\, \lambda\left(\bigcup_{k=0}^{n} I_\varepsilon\left(S_k(x)\right) \setminus \bigcup_{k=1}^{n} I_\varepsilon\left(S_k(x)\right)\right) \\ &=\, \lambda\left([-\varepsilon, \varepsilon] \setminus \bigcup_{k=1}^{n} I_\varepsilon\left(S_k(x)\right)\right) \,\geq\, 2\varepsilon\, \mathbf{1}_{\{\min_{k=1,\ldots,n} |S_k| > 2\varepsilon\}}(x)\,, \end{aligned} \tag{1.66}$$

where in the second equality we used that the Lebssgue measure λ is translation invariant. Note that from the last equality (in (1.66)) we see that $m_{n+1}^\varepsilon(x) - m_n^\varepsilon(\theta x)$ does not increase as n increases and therefore $\mathbf{E}[m_{n+1}^\varepsilon] - \mathbf{E}[m_n^\varepsilon]$ has the same property.

All this implies that

$$\Lambda\left(\left\{x:\min_{k=1,\dots,n}|S_k(x)|>2\varepsilon\right\}\right)\le\frac{1}{2\varepsilon}\left(\mathbf{E}\left[m^\varepsilon_{n+1}\right]-\mathbf{E}\left[m^\varepsilon_n\right]\right)$$
$$\le\frac{1}{2\varepsilon(n+1)}\sum_{k=0}^{n}\left(\mathbf{E}\left[m^\varepsilon_{k+1}\right]-\mathbf{E}\left[m^\varepsilon_k\right]\right)\le\frac{\mathbf{E}\left[m^\varepsilon_{n+1}\right]}{2\varepsilon(n+1)}\overset{n\to\infty}{\longrightarrow}0\,,\,,\tag{1.67}$$

that means

$$\Lambda\left(\left\{x:\min_{k\in\mathbb{N}}|S_k(x)|>2\varepsilon\right\}\right)=0\,.\tag{1.68}$$

Actually, $S_{k+m}-S_m=S_k\circ\theta^m$ and $S_k\circ\theta^m$ has the same law as S_k so for every $\varepsilon>0$ and every $m=0,1,\dots$

$$\Lambda\left(\left\{x:\min_{k\in\mathbb{N}}|S_{k+m}(x)-S_m(x)|>2\varepsilon\right\}\right)=0\,,\tag{1.69}$$

and therefore the Λ-measure of the event

$$E_0:=\{x:\,\exists\,\varepsilon>0\text{ and }m\in\mathbb{N}\cup\{0\}\text{ such that }|S_{k+m}(x)-S_m(x)|>2\varepsilon\ \forall\,k\in\mathbb{N}\}\,,\tag{1.70}$$

is zero and this is impossible because $\{x:\,S_n(x)\to\infty\}\subset E_0$. In fact if $S_n(x)\to\infty$ we can choose m equal to the last time n at which $S_n(x)$ is equal to its global minimum, i.e. $m=\max\{n=0,1,\dots:\,S_n(x)=\min_k S_k(x)\}$, and we can choose $\varepsilon=\inf_{k=1,2,\dots}|S_{m+k}(x)-S_m(x)|/3$, which is in fact a minimum and it is positive because of the way defined m and because $S_{m+k}(x)-S_m(x)\to\infty$ for $k\to\infty$. □

1.3.4. The Furstenberg Theorem and Formula: a proof. This section is devoted to the proof of Theorem 1.8. The arguments we present naturally give more results that we collect in the following statement:

PROPOSITION 1.17. *Under the hypotheses of Theorem 1.8*

(1) For every $\overline{x},\overline{y}\in\mathbb{B}$ with $\overline{x}\neq\overline{y}$ we have that $\mathbb{P}(\,\mathrm{d}\omega)$-a.s.

$$\lim_n\frac{1}{n}\log\mathbf{d}(Y_n(\omega)\dots Y_1(\omega)\cdot\overline{x},Y_n(\omega)\dots Y_1(\omega)\cdot\overline{y})=-2\gamma\,.\tag{1.71}$$

(2) There exists a random variable $\omega\mapsto V(\omega)$ taking values in $\mathbb{R}^2\setminus\{0\}$ such that $\mathbb{P}(\,\mathrm{d}\omega)$-a.s.

$$\lim_n\frac{1}{n}\log\|Y_n\dots Y_1\,V(\omega)\|=-\gamma\,,\tag{1.72}$$

and for every non zero random variable U which a.s. is not collinear with V (i.e., the probability that $U(\omega)$ and $V(\omega)$ are not on the same ray is one) we have

$$\lim_n\frac{1}{n}\log\|Y_n\dots Y_1\,U(\omega)\|=\gamma\,,\tag{1.73}$$

Proof of Theorem 1.8. For the existence of a unique μ-invariant probability see Remark 1.12 and ν is diffuse by Lemma 1.9.

We recall that $X_k = Y_k^*$ and in this proof $S_n := Y_n \dots Y_1$, so $S_n = M_n^*$ (with M_n used in Sec. 1.3.2). By Proposition 1.11 we know that a.s. $S_n^*(\omega)\nu^* \Rightarrow \delta_{\overline{Z}(\omega)}$ so, by Lemma 1.15, we have that for every $x \in \mathbb{R}^2$ a.s.

$$\lim_n \|S_n\| = \infty \quad \text{and} \quad \lim_n \frac{\|S_n(\omega)x\|}{\|S_n(\omega)\|} \,=\, \left|\left\langle \overline{Z}(\omega), x\right\rangle\right| . \tag{1.74}$$

Since the law of $\overline{Z}(\omega)$ is diffuse $\langle \overline{Z}(\omega), x\rangle \neq 0$ a.s. and this largely implies that $\log\|S_n(\omega)x\|$ and $\log\|S_n(\omega)\|$ are asymptotically equivalent and, by Theorem 1.5 that says that $(1/n)\log\|S_n(\omega)\| \to \gamma$, (1.36) is established.

For the integral formula (1.37) we recall that the relevant cocycle is $\sigma(M, \overline{x}) = \log(\|Mx\|/\|x\|)$. We have

$$\int_{\mathbb{B}}\int_{\mathbb{G}} \sigma_+(M, \overline{x})\mu(\,\mathrm{d}M)\nu(\,\mathrm{d}\overline{x}) \,\le\, \int_{\mathbb{G}} \log_+\|M\|\mu(\,\mathrm{d}M) \,=\, \mathbb{E}\left[\log_+\|Y_1\|\right] . \tag{1.75}$$

Moreover from (1.74) we have that

$$\lim_n \sigma\left(S_n, \overline{x}\right) \,=\, \infty\,, \tag{1.76}$$

if $\langle \overline{Z}(\omega), \overline{x}\rangle \neq 0$ and this happens $\mathbb{P}\otimes\nu(\,\mathrm{d}(\omega, \overline{x}))$-a.s. by the Fubini-Tonelli Theorem using either that ν is diffuse or that the law of $\overline{Z}$ is diffuse. We can therefore apply Lemma 1.16 and we obtain that $\sigma \in \mathbb{L}^1$ and

$$\int_{\mathbb{B}}\int_{\mathbb{G}} \log\frac{\|Mx\|}{\|x\|}\mu(\,\mathrm{d}M)\nu(\,\mathrm{d}\overline{x}) \,>0\,. \tag{1.77}$$

We are therefore done once we show that the left-hand side of inequality (1.77) coincides with γ. For this we remark that by Proposition 1.4 (strictly speaking, Proposition 1.4 is stated and proven with a different choice of space $\mathbb{B}$, but the proof is identical for the $\mathbb{B}$ we are using now) $\lim_n \sigma(S_n(\omega), \overline{x})/n =: \Phi(\omega, \overline{x})$, $\mathbb{P}\otimes\nu(\,\mathrm{d}(\omega, \overline{x}))$-a.s. and in $\mathbb{L}^1$. In particular $\int_{\mathbb{B}}\int_{\Omega}\Phi(\omega, \overline{x})\mathbb{P}(\,\mathrm{d}\omega)\nu(\,\mathrm{d}\overline{x}) = \int_{\mathbb{B}}\int_{\mathbb{G}}\sigma(M, \overline{x})\mu(\,\mathrm{d}M)\nu(\,\mathrm{d}\overline{x})$: note that this is the left-hand of (1.77). At this point we recall that we have just proven that $\log\|S_n(\omega)x\| \sim \log\|S_n(\omega)\| \sim n\gamma$ for $n\to\infty$ and of course $\sigma(S_n, \overline{x}) = \log(\|S_n(\omega)x\|/\|x\|) \sim \log\|S_n(\omega)x\|$. We therefore conclude that $\Phi(\omega, \overline{x}) = \gamma$, $\mathbb{P}\otimes\nu(\,\mathrm{d}(\omega, \overline{x}))$-a.s., which directly yields that the left-hand of (1.77) is equal to γ. □

Proof of Proposition 1.17. With the same notation of the previous proof, by (1.74), by recalling (1.35) and that $|\det(S_n(\omega)|=1$ we see that for $\overline{x}\neq\overline{y}$ that are not orthorgonal to $\overline{Z}(\omega)$

$$\mathbf{d}\left(S_n(\omega)\cdot\overline{x}, S_n(\omega)\cdot\overline{y}\right) \,=\, \frac{|\det(x,y)|}{\|S_n(\omega)x\|\|S_ny\|} \overset{n\to\infty}{\sim} \frac{|\det(x,y)|}{|\langle Z(\omega), x\rangle||\langle Z(\omega), y\rangle|}\frac{1}{\|S_n(\omega)\|^2}\,. \tag{1.78}$$

Since the law of $\overline{Z}(\omega)$ is diffuse we have that $\mathbb{P}(\langle Z, x\rangle\langle Z, y\rangle = 0) = 0$ so, by (1.36), we obtain (1.71).

We now move to (1.72) and we remark that (1.72) is strongly suggested by $\lim_n \|S_n(\omega)x\|/\|S_n(\omega)\| = |\langle Z(\omega), x\rangle|$, with $V(\omega)$ perpendicular to $Z(\omega)$. In fact $\|S_n(\omega)\|^2 = \|S_n^*(\omega)S_n(\omega)\|$ is just the largest of the two eigenvalues of $S_n^*(\omega)S_n(\omega)$. So the other eigenvalue is $1/\|S_n(\omega)\|^2$ and if we call $u_n(\omega)$ and $v_n(\omega)$ the two (normalized) eigenvectors, we have $\|S_n(\omega)u_n(\omega)\| = \|S_n(\omega)\|$ as well as $\|S_n(\omega)v_n(\omega)\| = 1/\|S_n(\omega)\|$. So we do have a direction $(v_n(\omega))$ along which the matrix product contracts exponentially with rate γ (and *therefore* all other directions expand with rate γ). Trouble is that $v_n(\omega)$ depends on n and we want to replace $v_n(\omega)$ with a vector that does not depend on n. Actually, this replacement is possible because $v_n(\omega)$ does converge to a limit unit vector and because it does so exponentially fast with a rate that is larger than γ (in fact, it is 2γ). This deep linear algebra statement is the $d = 2$ case of a celebrated result (Oseledets Theorem): statement and proof are in Appendix A (Theorem A.3).

The proof of Proposition 1.17 is therefore complete. □

We complete this section with the following result that is weaker than Proposition 1.17, but goes in the same direction and can be proven in a more straightforward way.

PROPOSITION 1.18. *Under the same hypotheses as for Theorem 1.8 and with $\overline{Z}$ as in* (1.74), $\mathbb{P}(\mathrm{d}\omega)$*-a.s. we have that*

$$\lim_{n\to\infty} \frac{S_n^*(\omega)S_n(\omega)}{\|S_n(\omega)\|^2} = z(\omega)z^*(\omega)\,, \tag{1.79}$$

where $z(\omega) \in \mathbb{R}^2$ belongs to $\overline{Z}(\omega)$ and $\|z(\omega)\| = 1$. Moreover, the two eigendirections $U_n(\omega) \in \mathbb{B}$ (of eigenvalue $\|S_n(\omega)\|^2$) and $V_n(\omega) \in \mathbb{B}$ (of eigenvalue $1/\|S_n(\omega)\|^2$) of $S_n^*(\omega)S_n(\omega)$ converge respectively to $\overline{Z}(\omega)$ and to $\overline{Z}(\omega)^\perp$.

PROOF. Let us consider the symmetric matrix $Q_n(\omega) := \frac{S_n^*(\omega)S_n(\omega)}{\|S_n(\omega)\|^2}$. For every $x \in \mathbb{R}^2$ we have (by the same argument like at the beginning of the proof of Theorem 1.8)

$$\langle x, Q_n(\omega)x\rangle = \frac{\|S_n(\omega)x\|^2}{\|S_n(\omega)\|^2} \overset{n\to\infty}{\longrightarrow} \langle z(\omega), x\rangle^2\,. \tag{1.80}$$

Therefore $Q_n(\omega)$ converges to the singular (symmetric) matrix $z(\omega)z^*(\omega)$. Note that the ray $\overline{Z}(\omega)$ is the range of $z(\omega)z^*(\omega)$, while $\overline{Z}(\omega)^\perp$ is the kernel of $z(\omega)z^*(\omega)$ The convergence of the eigendirections follows from the converge of the matrix. □

CHAPTER 2

On the disordered Ising model

2.1. The basic example: the one dimensional Ising chain

We consider the one dimensional Ising model with random external field, that is a probability (*Gibbs measure*) on the spin configurations $\sigma \in \{-1,+1\}^N$. We choose periodic boundary conditions for simplicity, that is $\sigma_{N+1} := \sigma_1$, but this can easily be generalized (see Remark 2.1). The Hamiltonian of this model is

$$H_N(\sigma) := -J\sum_{j=1}^N \sigma_j\sigma_{j+1} - \sum_{j=1}^N h_j\sigma_j\,, \tag{2.1}$$

where $J>0$ and (h_j) is the realization of a sequence of IID random variable: the law of (h_j) is denoted by $\mathbb{P}$, but we stress that we really view (h_j) as a given (frozen, quenched,...) realization. We then introduce the probability measure $\mu_{N,(h_j)}$ by setting

$$\mu_{N,(h_j)}(\sigma) := \frac{\exp\left(-H_N(\sigma)\right)}{Z_{N,(h_j)}}\,, \tag{2.2}$$

where $Z_{N,(h_j)}$ is the normalization and therefore

$$Z_{N,(h_j)} = \sum_{\sigma\in\{-1,1\}^N} \exp\left(-H_N(\sigma)\right)\,. \tag{2.3}$$

$Z_{N,(h_j)}$ is called *partition function* and it contains a lot of information about $\mu_{N,(h_j)}$. Notice in fact that

$$\partial_J \log Z_{N,(h_j)} = \sum_\sigma \left(\sum_{j=1}^N \sigma_j\sigma_{j+1}\right)\mu_{N,(h_j)}(\sigma) =: \mathbf{E}_{N,(h_j)}\left[\sum_{j=1}^N \sigma_j\sigma_{j+1}\right]\,, \tag{2.4}$$

which is the expected value of the Hamiltonian under $\mu_{N,(h_j)}$ and if $h=\mathbb{E}[h_1]$ then

$$\partial_h \frac{1}{N}\log Z_{N,(h_j)} = \mathbf{E}_{N,(h_j)}\left[\frac{1}{N}\sum_{j=1}^N \sigma_j\right]\,, \tag{2.5}$$

which is the expected value of the spatial average of the spins. Since we are interested in very large values of N we introduce the *free energy density*

$$\mathrm{F} = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_{N,(h_j)}\,. \tag{2.6}$$

We will see just below that this limit exists: in fact almost surely in the realization of (h_j)

$$\mathrm{F} \,=\, \lim_{N\to\infty} \frac{1}{N} \log Z_{N,(h_j)}\,, \tag{2.7}$$

and the convergence holds also in $\mathbb{L}^1$. The free energy density F is a function of J and of the law of (h_j). In particular, it may be viewed as a function of h and, for example, we do have that

$$\partial_h \mathrm{F} \,=\, \lim_N \mathbb{E}\mathbf{E}_{N,(h_j)}\left[\frac{1}{N}\sum_{j=1}^N \sigma_j\right]\,, \tag{2.8}$$

provided that $\partial_h\mathrm{F}$ exists (and, in this case, $\mathbb{E}$ can be removed from the right-hand side (2.8)). This result just follows from the fact that $h \mapsto \log Z_{N,(h_j)}$ is convex and convexity carries with itself also the information that $h \mapsto \mathrm{F}$ is continuous and C^1 except possibly for a countable number of values. We will see later on that F is in fact very regular.

Now we explain that $Z_{N,(h_j)}$ can be written as the trace of the product of the IID random matrices (called *transfer matrices*)

$$T_j \,:=\, \begin{pmatrix} e^{h_j} & 0 \\ 0 & e^{-h_j}\end{pmatrix}\begin{pmatrix} e^{J} & e^{-J} \\ e^{-J} & e^{J}\end{pmatrix} \,=\, \begin{pmatrix} e^{J+h_j} & e^{-J+h_j} \\ e^{-J-h_j} & e^{J-h_j}\end{pmatrix}\,. \tag{2.9}$$

In fact

$$\begin{aligned} Z_{N,(h_j)} \,&=\, \sum_{\sigma_1\in\{-1,+1\}}\sum_{\sigma_2\in\{-1,+1\}}\cdots\sum_{\sigma_N\in\{-1,+1\}}\prod_{j=1}^N e^{J\sigma_j\sigma_{j+1}+h_j\sigma_j} \\ &=\, \sum_{\sigma_1\in\{-1,+1\}} \left(T_1T_2\ldots T_N\right)_{k(\sigma_1),k(\sigma_1)} \,=\, \mathrm{trace}\left(T_1T_2\ldots T_N\right)\,, \end{aligned} \tag{2.10}$$

where we have used the periodic boundary conditions $\sigma_{N+1} = \sigma_1$ and we have introduced the function $k : \{-1, 1\} \to \{1, 2\}$ defined by $k(-1) = 2$ and $k(1) = 1$.

REMARK 2.1. *Other boundary conditions can be chosen: for example if we fix $\sigma_1 = 1$ and $\sigma_{N+1} = -1$ then the arising partition function would simply be equal to $(T_1T_2\ldots T_N)_{1,2}$.*

From this matrix product representation we can easily extract that the limits in (2.6) and (2.7) exist if $h_1 \in \mathbb{L}^1$.

PROPOSITION 2.2. *Assume $h_1 \in \mathbb{L}^1$. For every $J > 0$ the limits in* (2.6) *and* (2.7) *exist and they are equal to the top Lyapunov exponent of the product of the random matrices (T_j). In short, $\mathrm{F} = \gamma$.*

PROOF. Let us first observe a fact of independent interest (and valid in general for matrices with positive entries even if we give the argument for the specific case we consider): for $j, k \in \{1, 2\}$

$$\lim_n \frac{1}{n} \log(T_1 T_2 \dots T_n)_{j,k} \, = \, \gamma \, , \tag{2.11}$$

a.s. and in $\mathbb{L}^1$. This of course yields the result for the trace, i.e. the convergence results we are after.

The claim (2.11) follows directly from

$$\left| \log(T_1 T_2 \dots T_n)_{j,k} - \log \sum_{j,k} (T_2 \dots T_{n-1})_{j,k} \right| \le 2J + |h_1| + |h_n| \, . \tag{2.12}$$

This follows by observing that with $h = h_n$

$$\begin{pmatrix} e^{J+h} & e^{-J+h} \\ e^{-J-h} & e^{J-h} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \, = \, \begin{pmatrix} e^{J+h} \\ e^{-J-h} \end{pmatrix} \, = \, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{cases} \le \exp(J + |h|) \, , \\ \ge \exp(-J - |h|) \, , \end{cases} \tag{2.13}$$

where the inequalities are meant *componentwise*. Note that the same bounds hold if we replace $(1, 0)$ by $(0, 1)$. In the same way for $h = h_1$

$$(1, 0) \begin{pmatrix} e^{J+h} & e^{-J+h} \\ e^{-J-h} & e^{J-h} \end{pmatrix} \, = \, \left(e^{J+h}, e^{-J+h} \right) \, = \, (1, 1) \times \begin{cases} \le \exp(J + |h|) \\ \ge \exp(-J - |h|) \, , \end{cases} \tag{2.14}$$

and the same bound holds if we replace $(1, 0)$ by $(0, 1)$. Therefore the claim (2.11) is proven and the proof is complete. □

Next we show that Furstenberg Theorem (Theorem 1.8) applies and we exploit the formula for γ that it gives.

PROPOSITION 2.3. *Assume $h_1 \in \mathbb{L}^1$ and nontrivial. Consider the Markov chain $X = (X_j)_{j=0,1,\dots}$ on $(0, \infty)$ defined by the random iteration*

$$X_{j+1} \mapsto e^{-2h_{j+1}} \frac{e^{-2J} + X_j}{1 + X_j e^{-2J}} \, . \tag{2.15}$$

Then X has a unique invariant probability m_J and

$$\mathrm{F} \, = \, J + \mathbb{E}[h_1] + \int_{(0,\infty)} \log\left(1 + e^{-2J} x\right) m_J(\mathrm{d}x) \, . \tag{2.16}$$

PROOF. First of all notice that $\det(T_j) = e^{2J} - e^{-2J} > 0$ is deterministic. So the transfer matrices can be trivially reduced to unit determinant: this reduction will be relevant only in verifying the non compactness property, so we will mostly work with (T_n) rather than $(T_n / \sqrt{2\sinh(2J)})$.

The integrability property is evident. For the rest note that T_1 is a matrix with positive entries, so by the Perron-Frobenius Theorem the largest eigenvalue is real, positive and simple. Since the determinant is positive, also the other eigenvalue is positive. Let us note immediately that this implies the non compactness property, because it implies that $T_n/\sqrt{2\sinh(2J)}$ has one eigenvalue larger than one.

Let us turn to irreducibility. For what follows we use the notation $T_1 x^+ = \lambda_+ x^+$ and $T_1 x^- = \lambda_- x^-$, with $\lambda_+ > \lambda_- > 0$. In particular, the ray containing x^+ and the one containing x_- are fixed. Moreover for every $x \neq 0$ not collinear with x^- we have that $T_1^n x$ becomes asymptotically collinear to x^+. In order to prove irreducibility it is therefore sufficient to show that the eigenvector corresponding to the leading eigenvalue of T_1 (for short: leading eigenvector) is not collinear with the leading eigenvector of T_2 when $h_1 \neq h_2$.

One can show this by making explicit eigenvalues and eigenvectors, and by analyzing their dependence on the value of h. This is what we do now (see however Remark 2.4 for a quicker argument). First of all $\lambda_\pm$ are identified by $\lambda_+\lambda_- = e^{2J} - e^{-2J}$ and $\lambda_+ + \lambda_- = 2e^J \cosh(h)$. Therefore

$$\lambda_\pm \;=\; e^J(\cosh(h) \pm \sqrt{\cosh(h)^2 - 2e^{-2J}\sinh(2J)}\,, \tag{2.17}$$

and

$$\frac{x_2^+}{x_1^+} \;=\; \frac{1}{2}e^J\left(\left(1+e^{-2h}\right)\sqrt{1-\frac{1-e^{-4J}}{\cosh(h)^2}} + e^{-2h} - 1\right)\,. \tag{2.18}$$

It is straightforward to see that $\lim_{h\to-\infty} x_2^+/x_1^+ = \infty$ and $\lim_{h\to\infty} x_2^+/x_1^+ = 0$. It is also immediate to see that x_2^+/x_1^+ decreases as $h<0$ increases. The fact that x_2^+/x_1^+ is decreasing also for h positive becomes clear if we observe that the term between parentheses in (2.18) can be rewritten as

$$\left(1+e^{-2h}\right)\tanh(h)\sqrt{1+\frac{e^{-4J}}{\cosh(h)^2-1}} + e^{-2h} - 1\,, \tag{2.19}$$

which is zero for $J=\infty$ for every $h>0$. The claim then follows because the square root term is decreasing for $h>0$ increasing. Therefore the leading eigenvectors of T_1 and of T_2 are not the same and no union of lines is invariant under the action of both T_1 and T_2.

So we can apply Theorem 1.8 and we have in particular

$$\mathrm{F} \;=\; \int_{\mathbb{B}} \mathbb{E}\left[\log\frac{\|T_1 v\|}{\|v\|}\right]\nu(\,\mathrm{d}\overline{v}) \;>\; \frac{1}{2}\log\left(2\sinh(2J)\right)\,, \tag{2.20}$$

where ν is the unique invariant probability on the projective space and the lower bound is $\gamma > 0$ in our case (for which the determinant is not 1). Since the matrices we are considering map the first quadrant to itself, as well as the third quadrant to itself, and since ν is unique, ν is supported only on the set of the lines going through the first and third quadrants and it is more practical to talk about a measure on an angle $\theta \in [0,\pi/2]$. In order to understand the action of the matrix we observe that for $y>0$

$$\begin{pmatrix} e^{J+h} & e^{-J+h} \\ e^{-J-h} & e^{J-h}\end{pmatrix}\begin{pmatrix}1\\ y\end{pmatrix} \;=\; \begin{pmatrix} e^{J+h} + e^{-J+h}y \\ e^{-J-h} + e^{J-h}y\;,\end{pmatrix} \tag{2.21}$$

so the action of the matrix on $(0,\infty)$ is

$$y \mapsto \frac{e^{-J-h}+e^{J-h}y}{e^{J+h}+e^{-J+h}y} = e^{-2h}\frac{e^{-2J}+y}{1+e^{-2J}y} =: Z\frac{\varepsilon+y}{1+\varepsilon y} =: Zf_\varepsilon(y)\,, \tag{2.22}$$

where we have introduced the *standard* variables [**9, 4**]

$$\varepsilon = e^{-2J} \quad \text{and} \quad Z = e^{-2h}\,. \tag{2.23}$$

Of course we can can go back to the action on $(0,\pi/2)$ by using $\theta = \arctan(y)$. Note that we can consider $[0,\infty]$ and $[0,\pi]$ instead, but 0 and infinity are transient and are not accessible from $(0,\infty)$.

REMARK 2.4. *Note that* (2.22) *and the fact that* $f_\varepsilon(y) > 0$ *for every* $y > 0$ *directly yield that different values of* Z *yield different eigendirections.*

By (2.21) it is clear that the Markov chain on the projective space we are after is equivalent to the Markov chain on the positive real numbers given in the statement of Proposition 2.3 (see (2.15)). And by Theorem 1.8 we know that it has a unique invariant probability that we call m_ε. Of course we know also that m_ε is diffuse. Finally, from (2.20), we see that

$$\mathrm{F} = -\frac{1}{2}(\mathbb{E}[\log Z]+\log\varepsilon)+\frac{1}{2}\int_{(0,\infty)}\mathbb{E}\left[\log\frac{(1+\varepsilon y)^2+Z^2(\varepsilon+y)^2}{1+y^2}\right]m_\varepsilon(\,\mathrm{d}y)\,, \tag{2.24}$$

where we have used that in the new variables

$$\begin{pmatrix}1\\ y\end{pmatrix} \mapsto \frac{1}{\sqrt{\varepsilon Z}}\begin{pmatrix}1+\varepsilon y\\ Z(\varepsilon+y)\end{pmatrix}. \tag{2.25}$$

We can simplify (2.24) by observing that

$$\int_{(0,\infty)}\mathbb{E}\left[\log\frac{(1+\varepsilon y)^2+Z^2(\varepsilon+y)^2}{1+y^2}\right]m_\varepsilon(\,\mathrm{d}y) = \\ \int_{(0,\infty)}\log\frac{(1+\varepsilon y)^2}{1+y^2}m_\varepsilon(\,\mathrm{d}y)+\int_{(0,\infty)}\mathbb{E}\left[\log\left(1+Z^2f_\varepsilon^2(y)\right)\right]m_\varepsilon(\,\mathrm{d}y) \tag{2.26}$$

and, by stationarity, $Zf_\varepsilon(Y)$ has the same law as Y, where Y and Z are independent and the law of Y is m_ε. Therefore we can replace $Z^2f_\varepsilon^2(y)$ with y^2 in the last integral and we obtain (2.16). □

As we will discuss in some detail in Section 2.3, one does not expect anything particularly interesting in the one dimensional Ising model. But there is a *pseudo-critical* behavior as $J\to\infty$, i.e. $\varepsilon \searrow 0$ which is meaningful in applications (see also Section 2.2). This is the limit of strong interaction, for fixed disorder intensity.

2.2. More on disordered systems and matrix products

TO BE WRITTEN

2.3. The Derrida-Hilhorst singularity

We now choose to work with the variables $\varepsilon = \exp(-2J) \in (0,1)$ and $Z_j = \exp(-2h_j) \in (0,\infty)$ that we have already used in the proof of Proposition 2.3. So

$$T_j = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon Z_j & Z_j \end{pmatrix} . \tag{2.27}$$

We are always going to assume (at least) that $\log Z_1 \in \mathbb{L}^1$. We call $\gamma(\varepsilon)$ the Lyapunov function of this matrix product:

$$\gamma(\varepsilon) = \int \log(1+\varepsilon t)\, m_\varepsilon(\mathrm{d}t)\,, \tag{2.28}$$

which is directly extracted from (2.16).

The origin of the question leads to the hypothesis $\varepsilon > 0$ but of course we can consider the case $\varepsilon = 0$: this case fails to be irreducible. Note also that the case $\varepsilon < 0$ can also be considered: in fact

$$D(1,-1)^{-1} \begin{pmatrix} 1 & -\varepsilon \\ -\varepsilon Z_j & Z_j \end{pmatrix} D(1,-1) = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon Z_j & Z_j \end{pmatrix} \tag{2.29}$$

where $D(a,b)$ is the diagonal matrix with $D(a,b)_{1,1} = a$ and $D(a,b)_{2,2} = b$. This directly yields that $\gamma(\varepsilon) = \gamma(-\varepsilon)$.

Corollary B.6 tells us that $\varepsilon \mapsto \gamma(\varepsilon)$ is real analytic for $\varepsilon \in (0,1)$, so the same is true for $\varepsilon \in (-1,1) \setminus \{0\}$. What happens for $\varepsilon \to 0$?

We are now going to explain the 2-scale idea in [**9**] that leads to the prediction that if $\mathbb{E}[\log Z] < 0$ and $\mathbb{E}[Z] \in (1,\infty)$ then

$$\gamma(\varepsilon) \stackrel{\varepsilon\to 0}{\sim} C|\varepsilon|^{2\alpha}\,, \tag{2.30}$$

where C is a positive constant and $\alpha \in (0,1)$ is the unique positive solution to the equation $\mathbb{E}[Z^\beta] = 1$. Note in fact that $\beta \mapsto \mathbb{E}[Z^\beta]$ is a convex function. The assumption that $\mathbb{E}[Z]$ is finite is telling us that this function is bounded in $[0,1]$ (hence continuous and even smooth in $(0,1)$) and $\mathbb{E}[\log Z] < 0$ is its derivative form the right at zero. Since this function in zero takes value one, $\mathbb{E}[Z^\beta] = 1$ has a unique solution $\beta \in (0,1)$ that we call α.

The point is understanding the invariant measure of the Markov chain on $(0,\infty)$ defined by the random iteration (2.22): $X_{n+1} = Z_{n+1} f_\varepsilon(X_n)$. It is of course equivalent to study the Markov chains (X_n/ε) or (εX_n): we are going to consider the $\varepsilon \searrow 0$ limit of these two processes.

(1) The random iteration corresponding to (X_n/ε) is

$$x \mapsto Z\frac{1+x}{1+\varepsilon^2 x} \stackrel{\varepsilon\searrow 0}{\longrightarrow} Z(1+x)\,. \tag{2.31}$$

Therefore we introduce the Markov chain (X_n^{sr}), that focuses on the *short range* (i.e., near the origin), defined by $X_{n+1}^{\mathrm{sr}} = Z_{n+1}(1 + X_n^{\mathrm{sr}})$.

(2) The random iteration corresponding to (εX_n) is

$$x \mapsto Z\frac{\varepsilon^2 + x}{1+x} \stackrel{\varepsilon\searrow 0}{\longrightarrow} Z\frac{x}{1+x}\,, \tag{2.32}$$

and we introduce the Markov chain (X_n^{lr}), that focuses on the *long range* (i.e., the tail), defined by $X_{n+1}^{\mathrm{lr}} = Z_{n+1} X_n^{\mathrm{lr}}/(1+X_n^{\mathrm{lr}})$.

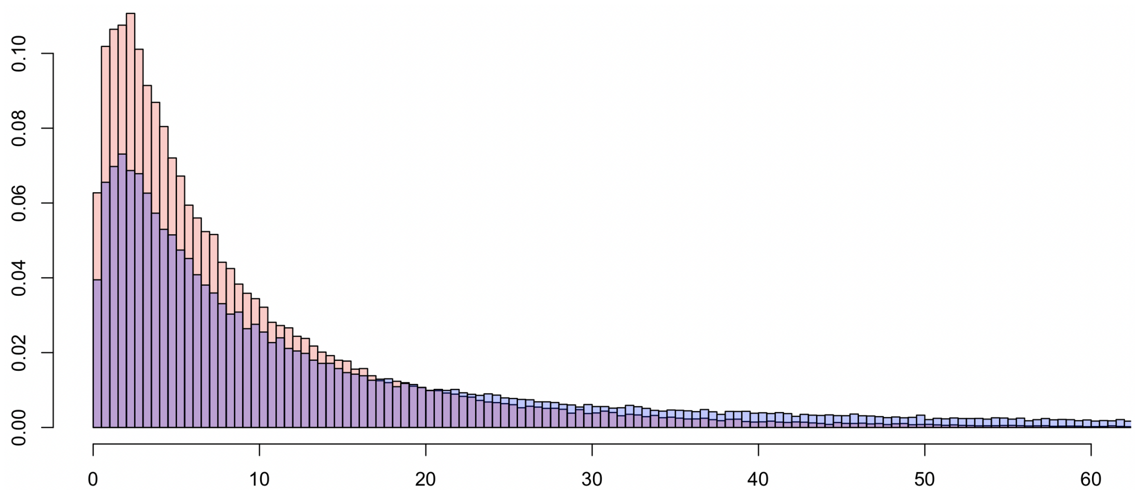

FIGURE 2.1. A numerical approximation of the invariant probability of the *short range* Markov chain (2.31), for $\varepsilon = 0.1$ (the more compact histogram) and $\varepsilon = 0$. The simulation uses $Z \sim U(0.05, 2.05)$, which yields $\mathbb{E}[\log Z] = -0.1893\ldots$ and $\alpha = 0.7369\ldots$ In fact, the histogram approaches the density of the invariant probability.

For the short range process we can even write an explicit formula: by iterating we see that

$$X_n^{\mathrm{sr}} \;=\; Z_n + Z_n Z_{n-1} + \ldots + Z_n Z_{n-1} \ldots Z_2 + Z_n Z_{n-1} \ldots Z_1 \left(1 + X_0^{\mathrm{sr}}\right) , \qquad (2.33)$$

where of course X_0^{sr} is independent of (Z_n). Then we remark that

$$X_n^{\mathrm{sr}} \;\sim\; Z_1 + Z_1 Z_2 + \ldots + Z_1 Z_2 \ldots Z_{n-1} + Z_1 Z_2 \ldots Z_n \left(1 + X_0^{\mathrm{sr}}\right) , \qquad (2.34)$$

and, if $\mathbb{E}[\log Z] < 0$, the right hand side converges a.s. to the random variable

$$Y \;:=\; \sum_{n=1}^{\infty} \prod_{j=1}^{n} Z_j \, , \qquad (2.35)$$

whose law ν_{sr} is therefore the invariant probability for the short range Markov chain. Since X_0^{sr} has an arbitrary law, this is the unique invariant probability of the chain. It is not completely straightforward to get quantitative properties on law of Y, but this random variable has been studied, notably because it come sup as fundamental tool in studying random walks in random environment. Notably, in [**20**, Th. 5] it is shown that there exists $c_{\mathrm{sr}} > 0$ such that

$$\nu_{\mathrm{sr}}((t, \infty)) \overset{t \to \infty}{\sim} \frac{c_{\mathrm{sr}}}{t^{\alpha}} \, , \qquad (2.36)$$

where ν_{sr} is the law of Y and (2.36) holds just assuming that the support of $\log Z$ is not contained in $c\mathbb{Z}$ for some $c > 0$ and that $\mathbb{E}[Z^{\alpha} \log_+ Z] < \infty$ (besides, of course, $\mathbb{E}[\log Z] < 0$ and $\mathbb{E}[Z^{\alpha}] = 1$). For the result in (2.36) we signal also [**3**].

The long range process has a very different nature. In fact it is not difficult to see that it is transient. This follows by observing that $X_{n+1}^{\mathrm{lr}} \le Z_{n+1} X_n^{\mathrm{lr}}$. Hence

$$X_n^{\mathrm{lr}} \le X_0 \prod_{j=1}^n Z_j = X_0 \exp\left(\sum_{j=1}^n \log Z_j\right), \tag{2.37}$$

so $\lim_n X_n^{\mathrm{lr}} = 0$ a.s.. As a matter of fact, the random walk $(\sum_{j=1}^n \log Z_j)$ is transient (and tends to $-\infty$). Therefore $(\log X_n^{\mathrm{lr}})$ is transient, and so is (X_n^{lr}). There is therefore no hope to find an invariant probability. Nevertheless one can find invariant measures: this problem has not been considered in full generality, but for example in [**17**] it is shown that an invariant measure ν_{lr} exists if

(1) the support of the law of Z is bounded and bounded away from zero;
(2) Z has C^1 density.

Moreover, under these conditions, ν_{lr} has a density (still denoted by $\nu_{\mathrm{lr}}(\cdot)$) and $\lim_{s\searrow 0} \nu_{\mathrm{lr}}(s)/s^{1+\alpha} = c > 0$. Since ν_{lr} is defined up to a multiplicative constant, the value of c is arbitrary. On the other hand, since we are assuming that the support of the law of Z is bounded, also the support of ν is bounded: it is straightforward to see that the support of ν_{lr} is bounded above by $\sup \mathrm{Supp}(Z)$. in fact, a more attentive analysis shows that $\sup \mathrm{Supp}(\nu_{\mathrm{lr}}) = \sup \mathrm{Supp}(Z) - 1$.

It is certainly possible to generalize this result to a wider class of distributions (less regular, without the conditions on the support) and obtain nevertheless that

$$\nu_{\mathrm{lr}}\left((s,\infty)\right) \stackrel{s\searrow 0}{\sim} \frac{c_{\mathrm{lr}}}{s^\alpha}, \tag{2.38}$$

where, again, $c_{\mathrm{lr}} > 0$ depends on an arbitrary choice. However such a result has been fully developed only in the restricted set-up explained above.

Remark 2.5. *It is not difficult to identify the supremum and the infimum of the support of m_ε, the invariant probability of our main Markov chain defined by the iteration* (2.22)*. When the support of Z is bounded away from* 0 *and* ∞*, it is just a matter of finding the (stable) fixed point in* $(0,\infty)$ *of* $y \mapsto z f_\varepsilon(y)$ *with* $z = z_- := \inf \mathrm{Supp}(Z) < 1$ *and* $z = z_+ := \sup \mathrm{Supp}(Z) > 1$*. Then*

$$\mathrm{Supp}\left(m_\varepsilon\right) \subset \left[\frac{\sqrt{4\varepsilon^2 z_- + (z_- - 1)^2} + z_- - 1}{2\varepsilon}, \frac{\sqrt{4\varepsilon^2 z_+ + (z_+ - 1)^2} + z_+ - 1}{2\varepsilon}\right], \tag{2.39}$$

so, to leading order in $\varepsilon \searrow 0$*, the support is between* ε *times* $2z_-/(1-z_-)$ *and* $1/\varepsilon$ *times* $z_+ - 1$*. These results are correct also if* $z_- = 0$ *and/or* $z_+ = \infty$*.*

Remark 2.6. *A way of grasping why the exponent* α *enters the game is that the "smoothing transformation"* $x \mapsto Zx$ *appears as large scale approximation of the short range transformation* (2.31) *and as short scale approximation of the long range transformation* (2.32)*. Passing to logarithm this is just the problem of finding positive measures that are invariant under convolution with the law of* $\log Z$*. One can then show that if* $\log Z$ *has a density with respect to the Lebesgue measure, then* *the only invariant measures are superpositions of measures with density proportional to* $t \mapsto \exp(-\alpha t)$ *with* $\alpha \in \mathbb{R}$ *solution to* $\mathbb{E}[Z^\alpha] = 1$. *Going back to* $(0, \infty)$ *by applying the exponential function we see that these densities become* $x \mapsto x^{-1-\alpha}$. *A priori one should consider also the presence of complex solutions* α, *but these are easily excluded because we are interested in positive measures.*

With ν_{sr} and ν_{lr} in our hands we can now define the probability measure g_ε on $(0, \infty)$. We use the notation $G_m(t) = m((t, \infty))$ for a positive measure m on $\mathbb{R}$ such that $\lim_{t\to\infty} m((t, \infty)) = 0$:

$$G_{g_\varepsilon}(t) = \begin{cases} G_{\nu_{\mathrm{sr}}}(t/\varepsilon) & \text{if } t \le 1\,, \\ a(\varepsilon) G_{\nu_{\mathrm{lr}}}(\varepsilon t) & \text{if } t > 1\,, \end{cases} \tag{2.40}$$

where $a(\varepsilon) := G_{\nu_{\mathrm{sr}}}(1/\varepsilon)/G_{\nu_{\mathrm{lr}}}(\varepsilon)$. Note that $\lim_{\varepsilon \searrow 0} a(\varepsilon)$ is equal to a positive constant (that depends on the choice we made for ν_{lr}) and, above all, that $G_{g_\varepsilon}(\cdot)$ is continuous.

We expect that g_ε is *close to* m_ε. Mathematically, the whole point is to show that this is true and in a sense that is sufficiently strong to control the error in the computation of the Lyapunov exponent. But let us proceed at a heuristic level assuming that (recall (2.28))

$$\gamma(\varepsilon) = \int \log(1 + \varepsilon t)\, m_\varepsilon(\,\mathrm{d}t) \overset{\varepsilon \searrow 0}{\sim} \int \log(1 + \varepsilon t)\, g_\varepsilon(\,\mathrm{d}t)\,. \tag{2.41}$$

And of course we can estimate precisely the rightmost term, by first re-expressing it via integration by parts:

$$\begin{aligned} \int \log(1 + \varepsilon t)\, g_\varepsilon(\,\mathrm{d}t) &= \varepsilon \int_0^\infty \frac{G_{g_\varepsilon}(t)}{1 + \varepsilon t}\,\mathrm{d}t \\ &= \varepsilon \int_0^1 \frac{G_{\nu_{\mathrm{sr}}}(t/\varepsilon)}{1 + \varepsilon t}\,\mathrm{d}t + \varepsilon \frac{G_{\nu_{\mathrm{sr}}}(1/\varepsilon)}{G_{\nu_{\mathrm{lr}}}(\varepsilon)} \int_1^\infty \frac{G_{\nu_{\mathrm{lr}}}(\varepsilon t)}{1 + \varepsilon t}\,\mathrm{d}t \\ &= \varepsilon^2 \int_0^{1/\varepsilon} \frac{G_{\nu_{\mathrm{sr}}}(t)}{1 + \varepsilon^2 t}\,\mathrm{d}t + \frac{G_{\nu_{\mathrm{sr}}}(1/\varepsilon)}{G_{\nu_{\mathrm{lr}}}(\varepsilon)} \int_\varepsilon^\infty \frac{G_{\nu_{\mathrm{lr}}}(t)}{1 + t}\,\mathrm{d}t\,, \end{aligned} \tag{2.42}$$

and, by (2.36), we readily see that the contribution of the first term is asymptotically equivalent to $(1-\alpha) c_{\nu_{\mathrm{sr}}} \varepsilon^{1+\alpha}$, but this sharp computation is useless because the second term is much larger: by (2.38)

$$\frac{G_{\nu_{\mathrm{sr}}}(1/\varepsilon)}{G_{\nu_{\mathrm{lr}}}(\varepsilon)} \int_\varepsilon^\infty \frac{G_{\nu_{\mathrm{lr}}}(t)}{1 + t}\,\mathrm{d}t \overset{\varepsilon \searrow 0}{\sim} \frac{c_{\mathrm{sr}}}{c_{\mathrm{lr}}} \varepsilon^{2\alpha} \int_0^\infty \frac{G_{\nu_{\mathrm{lr}}}(t)}{1 + t}\,\mathrm{d}t \sim c_{\mathrm{sr}} \varepsilon^{2\alpha} \int_0^\infty \log(1 + t) \frac{\nu_{\mathrm{lr}}}{c_{\mathrm{lr}}}(\,\mathrm{d}t)\,, \tag{2.43}$$

and we obtain (2.30) with a precise value of C which depends on c_{sr}, but luckily not on c_{lr}. It depends of course on the long range distribution via the integral appearing in the last term in (2.43): this integral is finite because of (2.36) and because the support of ν_{lr} is bounded if the support of the law of Z is bounded.

REMARK 2.7. *What happens when* $Z = \exp(-2h)$ *is non random? Computations are elementary but consequences are deep: it is more practical to go back to the* *notation of* (2.17) *and we find (just a Perron-Frobenius eigenvalue computation)*

$$\mathrm{F}(J,h) \;=\; J+\begin{cases}\log\left(\cosh(h)+|\sinh(h)|\sqrt{1+e^{-4J}/(\sinh(h))^2}\right) & \textit{if } h\neq 0\,,\\ \log\left(1+\exp(-2J)\right) & \textit{if } h=0\,.\end{cases} \tag{2.44}$$

and for $J\to\infty$

$$\mathrm{F}(J,h) \;=\; J+\begin{cases}|h|+\frac{e^{-|h|}}{2\sinh(|h|)}\exp(-4J)+O\left(\exp(-8J)\right) & \textit{if } h\neq 0\,,\\ \exp(-2J)+O\left(\exp(-4J)\right) & \textit{if } h=0\,.\end{cases} \tag{2.45}$$

It is also useful to remark that for $h\neq 0$

$$\partial_h \mathrm{F}(J,h) \;=\; \frac{\mathrm{sign}(h)}{\sqrt{1+e^{-2J}/\sinh(|h|)}} \;=\; \mathrm{sign}(h)\left(1-\frac{e^{-2J}}{2\sinh(|h|)}\right)+O\left(e^{-4J}\right)\,, \tag{2.46}$$

and since $\partial_h \mathrm{F}(J,h)$ *is the expectation of the mean of the spins, hence it is the expectation of each spin by translation invariance, this result makes clear that, for* J *large and* $h\neq 0$*, almost all spins are aligned with the magnetic fields (i.e., their sign coincide with the sign of the field). The density of the mistakes is proportional to* $\exp(-4J)=\varepsilon^2$ *and, by working a bit more, one can show that these deviations for* J *large are essentially isolated "mistakes".*

Things are different for $h=0$ *because a more unstable phenomenon sets in: the system is equally well on* $+1$ *or* -1*, so it will organize in long stretches of* $+1$ *and* -1 *for* J *large. This is a "domain-wall" structure and the walls have density* $\exp(-2J)$*: this is the signature of a "pseudo-phase coexistence". One can work this out in detail for* h *non random, but we rather signal that when* h*, i.e.* Z*, is random the result* (2.30) *strongly suggests that there is a complex "domain-wall" structure, connected to a "frustration phenomenon".*

2.4. Mathematics results on the Derrida-Hilhorst singularity

The arguments in the previous section provide a probability g_ε that should be close to the invariant probability we are after. In fact, an analogous 2-scale argument can be performed also for the $\mathbb{E}[\log Z]=0$ case (in a sense, this is the $\alpha=0$ case) and it leads to the prediction that

$$\gamma(\varepsilon) \overset{\varepsilon\to 0}{\sim} \frac{c}{\log(1/|\varepsilon|)}\,, \tag{2.47}$$

for a $c>0$ that is about as explicit as for the $\mathbb{E}[\log Z]<0$ case. Note that this case is more singular than the previous one and this should not come as a surprise: $\gamma(\varepsilon)$ is singular at the origin because there the matrix is not irreducible. But in the case $\alpha=0$ the behavior of the logarithm of the two terms on the diagonal coincides. Said otherwise, there is no separation of the two Lyapunov exponents in the $\varepsilon=0$ case if $\mathbb{E}[\log Z]=0$.

We now state a result under the hypothesis that the (nontrivial!) support of the law of Z is bounded and bounded away from zero. This means that $\mathbb{E}[Z^\beta]<\infty$ for every $\beta\in\mathbb{R}$. Then it is not difficult to see that α – unique non zero solution β to $\mathbb{E}Z^\beta=1$ – is well defined as soon as $\mathbb{P}(Z>1)>0$ and $\mathbb{P}(Z<1)>0$, with the provision that $\mathbb{E}[\log Z] \neq 0$. In fact if $\mathbb{E}[\log Z] = 0$ we have that $\mathbb{E}Z^\beta = 1$ if and only if $\beta = 0$. Moreover if $\mathbb{P}(Z > 1) = 0$ we see that $\mathbb{E}[Z^\beta] < 1$ for every $\beta > 0$ and $\mathbb{E}[Z^\beta] > 1$ for every $\beta > 0$: that is why we want to avoid these (trivial cases).

THEOREM 2.8 ([**17, 18**]). *We consider the top Lyapunov exponent associated to the sequence* (T_j)*, see* (2.27) *and we assume that*

(1) the support of the law of Z *is bounded and bounded away from zero;*

(2) the law of Z *has a* C^1 *density;*

then if $\alpha \in (0,1)$ *there exist* $C > 0$ *and* $\delta > 0$ *such that*

$$\gamma(\varepsilon) \,=\, C|\varepsilon|^{2\alpha} + O\left(|\varepsilon|^{2\alpha+\delta}\right)\,, \tag{2.48}$$

and if $\alpha = 0$ *there exist* $C_1 > 0$*,* $C_2 \in \mathbb{R}$ *and* $\delta > 0$ *such that*

$$\gamma(\varepsilon) \,=\, \frac{C_1}{C_2 - \log(|\varepsilon|)} + O\left(|\varepsilon|^{\delta}\right)\,. \tag{2.49}$$

There is no loss of generality in assuming $\mathbb{E}[\log Z] < 0$. In fact

$$\begin{pmatrix} 1 & \varepsilon \\ \varepsilon Z & Z \end{pmatrix} \,=\, Z\begin{pmatrix} 1/Z & \varepsilon/Z \\ \varepsilon & 1 \end{pmatrix} \,=\, Z\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & \varepsilon \\ \varepsilon/Z & 1/Z \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.50}$$

so the Lyapunov exponent when we replace Z with $1/Z$ is equal to $-\mathbb{E}[\log Z] + \gamma(\varepsilon)$. But of course the hypothesis that $\mathbb{E}[Z] > 1$ is a real restriction. It is claimed in [**9**] that if $\mathbb{E}[Z^n] < 1$ but $\mathbb{E}[Z^{n+1}] > 1$ than we should have

$$\gamma(\varepsilon) \,=\, c_1\varepsilon^2 + \ldots + c_n\varepsilon^{2n} + C\varepsilon^{2\alpha} + o\left(\varepsilon^{2\alpha}\right)\,, \tag{2.51}$$

with explicit $c_1, \ldots, c_n \in \mathbb{R}$ and an implicit $C > 0$. Results in this direction can be found in [**13**], but they are still far from identifying the $C\varepsilon^{2\alpha}$ term in (2.51).

Of course another limit of the result in Theorem 2.8 is the hypothesis on the support of the law of Z and of the existence of a C^1 density (the result for $\alpha = 0$ actually holds under weaker conditions: in [**18**] the support condition is removed, and the density needs only to be Hölder continuous in an appropriate uniform way). It is not known how much these assumptions can be relaxed: in [**9**] an exactly solvable case is given (with Z that takes only two values, one of which is zero) in which $\gamma(\varepsilon) \sim H(\log(1/\varepsilon))\varepsilon^{2\alpha}$, with $H(\cdot)$ a non trivial periodic function. It is possible that this log-periodic amplitude persists when Z is supported contained in $\{x^n : n \in \mathbb{Z}\}$ for a given $x > 0$ and it may be that Theorem 2.8 can be generalized out of these cases.

We complete this section by giving some hints about how the proof of Theorem 2.8 goes.

Sketch of proof of Theorem 2.8. Let us discuss first the case $\mathbb{E}[\log Z] < 0$, i.e. $\alpha \in (0,1)$. There is a first step: show existence and establish asymptotic properties of both ν_{sr} and ν_{lr}. The importance of the asymptotic properties are already clear because of the computation (2.43), but sharper asymptotic control is needed for the rest of the proof. Once this step is performed we have in our hands the probability

g_ε and we need to show that it is sufficiently close to the invariant probability m_ε. For this we introduce a *norm* of differences of probabilities on $(0, \infty)$: for $\beta \in (0\alpha)$

$$\|\|\nu_1 - \nu_2\|\|_\beta := \int_0^\infty x^{1-\beta} |G_{\nu_1}(x) - G_{\nu_2}(x)| \, \mathrm{d}x \,. \tag{2.52}$$

One can check (without much effort) that if define $L_\varepsilon[\nu] = \int \log(1 + \varepsilon^2 t)\nu(\,\mathrm{d}t)$ we have

$$|L_\varepsilon(\nu_1) - L_\varepsilon(\nu_2)| \le \varepsilon^{2\beta} \|\|\nu_1 - \nu_2\|\|_\beta \,, \tag{2.53}$$

and if we call T_ε the one step Markov transition matrix for Markov process (let us choose the short range one, i.e. (2.31), but this is an arbitrary choice: with this choice and if we call $\widetilde{m}_\varepsilon$ the invariant probability on this scale, i.e. $G_{\widetilde{m}_\varepsilon}(t) = G_{m_\varepsilon}(t\varepsilon)$, then $\gamma(\varepsilon) = L_\varepsilon(\widetilde{m}_\varepsilon)$) then for every ν_1 and ν_2

$$\|\|T_\varepsilon \nu_1 - T_\varepsilon \nu_2\|\|_\beta \le \mathbb{E}[Z^\beta] \, \|\|\nu_1 - \nu_2\|\|_\beta \,, \tag{2.54}$$

and remark that $\mathbb{E}[Z^\beta] < 1$. This contraction property can be very useful because

$$\|\|\widetilde{m}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta = \|\|T_\varepsilon \widetilde{m}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta \le \|\|T_\varepsilon \widetilde{m}_\varepsilon - T_\varepsilon \widetilde{g}_\varepsilon\|\|_\beta + \|\|T_\varepsilon \widetilde{g}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta \,, \tag{2.55}$$

where, once again, $G_{\widetilde{g}_\varepsilon}(t) = G_{g_\varepsilon}(t\varepsilon)$. By applying the contraction property (2.54) to the first addendum in the rightmost term in (2.4) we readily obtain that

$$\|\|\widetilde{m}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta \le c_\beta \|\|T_\varepsilon \widetilde{g}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta \,, \tag{2.56}$$

with $c_\beta := 1/(1 - \mathbb{E}[Z^\beta])$. The inequality (2.56) is very useful because the right-hand side contains only $\widetilde{g}_\varepsilon$, a probability on which we have full control. We are therefore left with controlling $\|\|T_\varepsilon \widetilde{g}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta$. Unfortunately (for these notes) such an estimate is quite heavy, and it depends on the adequately sharp control on the asymptotic behaviors of ν_{sr} and ν_{lr} that we mentioned above. But it should be clear that it suffices to show that $\|\|T_\varepsilon \widetilde{g}_\varepsilon - \widetilde{g}_\varepsilon\|\|_\beta = o(\varepsilon^{2(\alpha - \beta)})$, and we can play on choosing $\beta < \alpha$ close to α.

If we want to treat the case $\alpha = 0$, there is no room to choose $\beta < \alpha$. But one can work with $\|\|\cdot\|\|_0$ and (2.54) still holds. But $\mathbb{E}[Z^0] = 1$ and there is no contraction anymore. The point that saves the game is that, by looking more closely, we can show that there is still a *micro*-contraction and one can establish a version of (2.56) with $c_\beta = c_0 = \infty$ replaced by an ε dependent term that tends to ∞ as $\varepsilon \searrow 0$. But it turns out that the control one can get on $\|\|T_\varepsilon \widetilde{g}_\varepsilon - \widetilde{g}_\varepsilon\|\|_0$ is sufficient to counter the effect of the diverging pre-factor.

"$\square$"

2.5. Continuum limits

We now consider an approximation of the random matrix product under analysis. It is a diffusion limit in which one can compute the invariant probability, hence the Lyapunov exponent becomes explicit (in terms of special functions). This approach has been initiated by [**15**] and was first applied in the Ising context by [**23**]: it is a *weak disorder approach*. The idea is simple: instead of (2.27) consider for $\Delta > 0$

$$T_j^\Delta = \begin{pmatrix} 1 & \varepsilon\Delta \\ \varepsilon\Delta Z_j^\Delta & Z_j^\Delta \end{pmatrix}, \tag{2.57}$$

where

$$Z_j^\Delta := \exp\left(\sigma\sqrt{\Delta}\eta_j - \alpha\frac{\sigma^2}{2}\Delta\right), \tag{2.58}$$

with $\sigma > 0$ and (η_j) are IID standard Gaussian variables (this choice has been made for simplicity, but it can be greatly generalized, see Remark 2.13). Note that

(1) $\mathbb{E}[\log Z_1^\Delta] = -\Delta\alpha\sigma^2/2$ and $\mathbb{E}[(Z_1^\Delta)\alpha] = 1$;
(2) as $\Delta \searrow 0$ we have that T_j^Δ tends to the identity matrix.

Point (2) says in particular that the dynamics induced by this matrix product is very slow, while (1) says that α corresponds to the same parameter for the case (2.27) (and if we set $\Delta = 1$ in (2.57) we just find a particular case of (2.27)).

We will now show that the dynamics happens on the time scale $1/\Delta$. For this we introduce the discrete time stochastic process $\{(X_1^\Delta(n), X_2^\Delta(n))\}_{n=0,1,\ldots}$ defined recursively from the deterministic initial condition $(X_1^\Delta(0), X_2^\Delta(0)) = (X_1(0), X_2(0))$ by

$$\begin{cases} X_1^\Delta(n+1) = X_1^\Delta(n) + \varepsilon X_2^\Delta(n)\Delta\,, \\ X_2^\Delta(n+1) = e^{\sigma\sqrt{\Delta}\eta_{n+1} - \alpha\frac{\sigma^2}{2}\Delta}\left(X_2^\Delta(n) + \varepsilon X_1^\Delta(n)\Delta\right), \end{cases} \tag{2.59}$$

which can be rewritten in a more compact fashion as

$$X^\Delta(n+1) = X^\Delta(n) + A^\Delta(n+1)X^\Delta(n)\,, \tag{2.60}$$

where

$$X^\Delta = \begin{pmatrix} X_1^\Delta \\ X_2^\Delta \end{pmatrix}, \qquad A^\Delta(n) = \begin{pmatrix} 0 & \varepsilon\Delta \\ \varepsilon\Delta Z^\Delta(n) & Z^\Delta(n) - 1 \end{pmatrix}. \tag{2.61}$$

So, $X^\Delta(n)$ results form the product of n independent matrices of the form $I + A^\Delta$.

Rewriting (2.59) as

$$\begin{cases} X_1^\Delta(n+1) - X_1^\Delta(n) = \varepsilon X_2^\Delta(n)\Delta\,, \\ X_2^\Delta(n+1) - X_2^\Delta(n) = \left(e^{\sigma\sqrt{\Delta}\eta_{n+1} - \alpha\frac{\sigma^2}{2}\Delta} - 1\right) X_2^\Delta(n) + e^{\sigma\sqrt{\Delta}\eta_{n+1} - \alpha\frac{\sigma^2}{2}\Delta}\varepsilon X_1^\Delta(n)\Delta, \end{cases} \tag{2.62}$$

and remarking that

$$e^{\sigma\sqrt{\Delta}\eta_{n+1} - \alpha\frac{\sigma^2}{2}\Delta} = 1 + \sigma\sqrt{\Delta}\eta_{n+1} - \alpha\frac{\sigma^2}{2}\Delta + \frac{\sigma^2}{2}\Delta\left(\eta_{n+1}\right)^2 + \ldots\,, \tag{2.63}$$

makes the following result plausible:

Proposition 2.9. *Consider the random process*

$$\left\{\left(X_1^\Delta\left(\lfloor t/\Delta\rfloor\right), X_2^\Delta\left(\lfloor t/\Delta\rfloor\right)\right)\right\}_{t\in[0,\infty)}, \tag{2.64}$$

with trajectories in the Skorokhod space $D([0,\infty),(0,\infty)^2)$*. As* $\Delta \searrow 0$ *we have that this process converges in law to the diffusion* $(X_1(\cdot), X_2(\cdot))$ *solution of the Itô stochastic system*

$$\begin{cases} \mathrm{d}X_1(t) \;=\; \varepsilon X_2(t)\,\mathrm{d}t\,, \\ \mathrm{d}X_2(t) \;=\; \left(\varepsilon X_1(t) + \frac{(1-\alpha)\sigma^2}{2} X_2(t)\right)\,\mathrm{d}t + \sigma X_2(t)\,\mathrm{d}B_t\,, \end{cases} \tag{2.65}$$

where $B.$ *is a standard Brownian motion and the initial condition is the same as for the* $\Delta > 0$ *case.*

The choice fo the space of CADLAG trajectories $D([0,\infty),(0,\infty)^2)$ has been made to accommodate the discrete trajectories, but one can easily define the $\Delta > 0$ process by affine interpolation and work in $C^0([0,\infty),(0,\infty)^2)$.

It natural to define also for $(X_1(\cdot), X_2(\cdot))$ a Lyapunov exponent and it turns out that this Lyapunov exponent can be made explicit: for this we recall one of the definitions of the modified Bessel function of 2^{nd} kind of index $\alpha \in \mathbb{C}$ and argument $x > 0$

$$K_\alpha(x) \;:=\; \int_0^\infty \exp\left(-x\cosh(t)\right)\cosh(\alpha t)\;\mathrm{d}t \;=\; \frac{1}{2}\int_0^\infty \frac{1}{y^{1+\alpha}}\exp\left(-\frac{x}{2}\left(y+\frac{1}{y}\right)\right)\,\mathrm{d}y\,. \tag{2.66}$$

Proposition 2.10. *For every* $\varepsilon \neq 0$ *and every* $(X_1(0), X_2(0)) \in \mathbb{R}\setminus\{(0,0)\}$ *the limit*

$$\lim_{t\to\infty}\frac{1}{t}\mathbb{E}\log\|(X_1(t),X_2(t))\| \;=:\; \gamma^0_{\sigma,\alpha}(\varepsilon)\,, \tag{2.67}$$

exists and does not depend on $(X_1(0), X_2(0))$*. Moreover*

$$\gamma^0_{\sigma,\alpha}(\varepsilon) \;=\; \frac{\sigma^2}{4}\left(\frac{xK_{\alpha-1}(x)}{K_\alpha(x)}\right)\,, \qquad \text{with} \quad x := \frac{4\varepsilon}{\sigma^2}\,. \tag{2.68}$$

How is this Lyapunov exponent $\gamma^0_{\sigma,\alpha}(\varepsilon)$ related to the Lyapunov exponent of the matrix product? We call $\gamma^\Delta(\varepsilon)$ the Lyapunov exponent associated to (T^Δ_j) defined in (2.57) and, by Theorem 1.8, we have that a.s.

$$\gamma^\Delta(\varepsilon) \;=\; \lim_{n\to\infty}\frac{1}{n}\log\left\|X^\Delta(n)\right\|\,. \tag{2.69}$$

The link between $\gamma^\Delta(\varepsilon)$ and $\gamma^0_{\sigma,\alpha}(\varepsilon)$ is a priori not clear. Recall in fact that in Proposition 2.9 the convergence is in $D([0,\infty),(0,\infty)^2)$ and this is equivalent to claiming convergence in $D([0,t],(0,\infty)^2)$, for every $t > 0$. So, in reality, the convergence is local in time or, at least, there is no uniformity in time. In this sense, the following result is not a priori obvious:

Proposition 2.11. *For every* $\varepsilon \neq 0$

$$\lim_{\Delta\searrow 0}\frac{\gamma^\Delta(\varepsilon)}{\Delta} \;=\; \gamma^0_{\sigma,\alpha}(\varepsilon)\,. \tag{2.70}$$

It will not come as a surprise now that one can analyze in great detail the $\varepsilon \to 0$ behavior of $\gamma^0_{\sigma,\alpha}(\varepsilon)$. What is possibly a bit surprising is that this continuum case yields behaviors that are in full agreement with the discrete case. And, in the continuum case, the results go much farther.

In the next statement $\Gamma(\cdot)$ denotes the Gamma function, see for example [**24**, 5.2] for definitions and properties.

Proposition 2.12. *Recall that $x = 4\varepsilon/\sigma^2$. For $\alpha \in (0,\infty) \setminus \mathbb{Z}$ we have for $\varepsilon \searrow 0$*

$$\frac{4}{\sigma^2}\gamma^0_{\sigma,\alpha}(\varepsilon) \;=\; c_1(\alpha)x^2 + \ldots + c_{\lfloor\alpha\rfloor}(\alpha)x^{2\lfloor\alpha\rfloor} + 2\frac{\Gamma(1-\alpha)}{\Gamma(\alpha)}\left(\frac{x}{2}\right)^{2\alpha} + O\left(x^{\min(2\lceil\alpha\rceil,4\alpha)}\right), \tag{2.71}$$

where $c_j(\cdot)$ is an explicit rational function.

For $\alpha \in \{1,2,\ldots\}$ we have

$$\frac{4}{\sigma^2}\gamma^0_{\sigma,\alpha}(\varepsilon) \;=\; c_1(\alpha)x^2 + \ldots + c_{\alpha-1}(\alpha)x^{2(\alpha-1)} + (-1)^\alpha\frac{2^{2-2\alpha}}{((\alpha-1)!)^2}x^{2\alpha}\log x + O\left(x^{2\alpha}\right), \tag{2.72}$$

where $c_j(\cdot)$ is the same rational function as in the non integer α case.

For $\alpha = 0$ we have

$$\gamma^0_{\sigma,0}(\varepsilon) \;=\; \frac{\sigma^2}{4\left(\log(1/x) - \gamma + \log 2\right)} + O\left(x^2\right), \tag{2.73}$$

with $\gamma = 0.577\ldots$ the Euler-Mascheroni constant.

Moreover the result for $\alpha < 0$ is directly recovered from (2.71)-(2.72) *by using the identity*

$$\frac{4}{\sigma^2}\gamma^0_{\sigma,\alpha}(\varepsilon) \stackrel{\alpha\le 0}{=} 2|\alpha| + \frac{4}{\sigma^2}\gamma^0_{\sigma,|\alpha|}(\varepsilon)\,. \tag{2.74}$$

The identity (2.74) is a simple consequence of the Bessel identity

$$xK_{1+\alpha}(x) \;=\; 2\alpha K_\alpha(x) + xK_{-1+\alpha}(x)\,, \tag{2.75}$$

that follows from (2.66) by integration by parts and by using the identity $K_\alpha(x) = K_{-\alpha}(x)$ which follows by the change of variable $y \to 1/y$ in (2.66).

2.6. Proofs

Proof of Proposition 2.9. Several works are dedicated to diffusion approximations. Here we exploit [**27**, pp. 266–272], notably [**27**, Assumptions (2.4)-(2.6), Theorem 11.2.3]. Alternatively, one can resort to [**10**, Corollary 4.2 in Chapter 7]. It is sufficient to check three hypotheses that we give in our set-up: recall (2.59)-(2.61)

- compute the local drift at $x \in \mathbb{R}^2$: uniformly for $x = (x_1, x_2)$ in compact sets

$$\begin{aligned} b^{\Delta}(x) \;=\; \frac{\mathbb{E}A^{\Delta}x}{\Delta} \;&=\; \begin{pmatrix} 0 & \varepsilon \\ \varepsilon\mathbb{E}\left[Z^{\Delta}\right] & \frac{\mathbb{E}\left[Z^{\Delta}-1\right]}{\Delta} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &\overset{\Delta\searrow 0}{\longrightarrow}\; b(x) \;:=\; b\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \text{with } b := \begin{pmatrix} 0 & \varepsilon \\ \varepsilon & (1-\alpha)\frac{\sigma^2}{2} \end{pmatrix}; \end{aligned} \tag{2.76}$$

- compute the diffusion matrix at x: again uniformly in x in compact subsets of $\mathbb{R}^2$

$$a^{\Delta}(x) \;:=\; \Delta^{-1}\mathbb{E}\left[A^{\Delta}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(x_1, x_2)(A^{\Delta})^{*}\right] \overset{\Delta\searrow 0}{\longrightarrow} a(x) \;:=\; \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 x_2^2 \end{pmatrix}; \tag{2.77}$$

- $\Delta^{-1}\mathbb{P}(\|A^{\Delta}\| \geqslant c) \to 0$ for every $c > 0$.

Then, since the stochastic differential system with drift $b(\cdot)$ and diffusion matrix $a(\cdot)$ has unique (strong) solution, the Markov chain X^{Δ} converges in law to the diffusion process with drift $b(\cdot)$ and diffusion matrix $a(\cdot)$, which is precisely the solution X to the stochastic differential system (2.65). This completes the proof of Proposition 2.9. □

Remark 2.13. *From the proof one readily sees that Proposition 2.9 holds well beyond the Gaussian case. In fact it suffices to consider a family of positive random variables* $\{Z^{\Delta}\}_{\Delta\in(0,\Delta_0)}$ *such that for some* $\sigma > 0$ *and* $\alpha \in \mathbb{R}$ *we have*

$$\lim_{\Delta\searrow 0}\frac{\mathbb{E}\left[Z^{\Delta}-1\right]}{\Delta} \;=\; \frac{1}{2}\sigma^2(1-\alpha) \qquad \text{and} \qquad \lim_{\Delta\searrow 0}\frac{\mathbb{E}\left[\left(Z^{\Delta}-1\right)^2\right]}{\Delta} \;=\; \sigma^2, \tag{2.78}$$

and such that for every $c > 0$

$$\lim_{\Delta\searrow 0}\frac{1}{\Delta}\mathbb{P}\left(\left|Z^{\Delta}-1\right| \,>\, c\right) \;=\; 0. \tag{2.79}$$

In addition, we can generalize Proposition 2.11 to the set up in this remark if we assume also that

$$\limsup_{\Delta\searrow 0}\left|\frac{\mathbb{E}[1/Z^{\Delta}]-1}{\Delta}\right| \;<\; \infty. \tag{2.80}$$

Proof of Proposition 2.10. We use the short-cut notation $\delta := \sigma^2(1-\alpha)/2 \in \mathbb{R}$ and work with $\varepsilon > 0$ without loss of generality. We start by showing that the process does not hit $(0,0)$. Recall that $(X_1(0), X_2(0)) \neq (0,0)$ and set $\tau_{(0,0)} := \inf\{t > 0 : (X_1(t), X_2(t)) = (0,0)\}$. For this let us consider $R(t) :=$

$\sqrt{X_1^2(t)+X_2^2(t)}$. By Itô's formula:

$$\begin{aligned} \mathrm{d}R(t) &= \frac{X_1}{R}\,\mathrm{d}X_1 + \frac{X_2}{R}\,\mathrm{d}X_2 + \frac{1}{2}\frac{X_1^2}{R^3}\,\mathrm{d}\langle X_2, X_2\rangle \\ &= \left(2\varepsilon\frac{X_1X_2}{R} + \delta\frac{X_2^2}{R} + \frac{\sigma^2}{2}\frac{X_1^2X_2^2}{R^3}\right)\mathrm{d}t + \sigma\frac{X_2^2}{R}\,\mathrm{d}B_t \\ &= R\left(2\varepsilon\frac{Y}{1+Y^2} + \delta\frac{Y^2}{1+Y^2} + \frac{\sigma^2}{2}\frac{Y^2}{(1+Y^2)^2}\right)\mathrm{d}t + R\left(\sigma\frac{Y^2}{1+Y^2}\right)\mathrm{d}B_t \\ &=: R\,D\,\mathrm{d}t + R\,Q\,\mathrm{d}B_t\,, \end{aligned} \tag{2.81}$$

where $Y := X_2/X_1 \in [-\infty,\infty]$ and $D = D(t)$ and $Q = Q(t)$ are uniformly bounded continuous stochastic processes ($\|D\|_\infty \le 2\varepsilon + |\delta| + \sigma^2/2$ and $Q \in [0,\sigma]$), defined up to $\tau_{(0,0)}$. Since, again by Itô's formula, we have

$$\mathrm{d}\log R(t) = \left(D(t) - \frac{1}{2}Q^2(t)\right)\mathrm{d}t + Q(t)\,\mathrm{d}B_t\,, \tag{2.82}$$

we see that $R(t)/R(0)$ is bounded away from zero on every compact time interval. This readily yields a contradiction if $\mathbb{P}(\tau_{(0,0)} < \infty) > 0$. Hence $\mathbb{P}(\tau_{(0,0)} < \infty) = 0$ and we have proven that the process does not hit the origin.

Next we work under the assumption that both $X_1(0) > 0$ and $X_2(0) > 0$ and that the diffusion does not hit the boundary of the quadrant in finite time: of course this covers also the case $X_1(0) < 0$ and $X_2(0) < 0$ and for all the other cases it suffices to show that the process does enter the (interior of the) first or third quadrant in a (random) time that is in $\mathbb{L}^1$.

By Itô formula for $Y = X_2/X_1$ we obtain

$$\mathrm{d}Y = \left(\varepsilon\left(1-Y^2\right) + \delta\,Y\right)\mathrm{d}t + \sigma Y\,\mathrm{d}B_t\,, \tag{2.83}$$

and we want to identify the invariant probability of (2.83) and show that *there is only one*. Actually, for uniqueness we have to restrict to measures supported on the first quadrant, but the trivial lack of uniqueness that is inherent to our problem is irrelevant. But let us first show that there is one invariant probability.

For this we observe that the generator of the evolution (2.83) acts on C^2 functions $f : (0,\infty) \to \mathbb{R}$ as

$$L_\varepsilon f(y) = \left(\varepsilon(1-y^2) + \delta\,y\right)f'(y) + \frac{\sigma^2}{2}y^2 f''(y) = \frac{\sigma^2}{2p_\varepsilon(y)}\left(y^2 p_\varepsilon(y) f'(y)\right)'\,, \tag{2.84}$$

where $p_\varepsilon(\cdot)$ is the probability density

$$p_\varepsilon(y) = \frac{1}{C_\varepsilon y^{1+\alpha}}\exp\left(-\frac{2\varepsilon}{\sigma^2}\left(y + \frac{1}{y}\right)\right) \qquad \text{with } C_\varepsilon = 2K_\alpha\left(4\varepsilon/\sigma^2\right)\,, \tag{2.85}$$

and $K_\alpha(\cdot)$ is defined in (2.66). This makes evident the reversible nature of the diffusion Y and, in particular, that $p_\varepsilon(\cdot)$ is an invariant probability density. But the transformation $S(t) := \log Y(t)$ makes things even more straightforward: S is a diffusion on $\mathbb{R}$ with additive noise and strongly confining potential:

$$\mathrm{d}S \;=\; -U'(S)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t \quad \text{ with } \; U(s) \;:=\; \varepsilon\left(\exp(-s) + \exp(s) + \frac{\alpha\sigma^2}{2\varepsilon}s\right)\,. \tag{2.86}$$

An invariant probability of this diffusion is $\widetilde{p}_\varepsilon(s) \propto \exp(-2U(s)/\sigma^2)$ and the generator has the familiar symmetric form $\widetilde{L}_\varepsilon g = (\sigma^2/2)(\widetilde{p}_\varepsilon g')'/\widetilde{p}_\varepsilon$, for $g \in C^2(\mathbb{R},\mathbb{R})$, see for example [**14**, p.111], to which one can refer also for the ergodic properties of the process and for the uniqueness of the invariant measure. We cite here one of the first works dealing with this issue [**21**] and were one can find also the (Pointwise) Ergodic Theorem we are going to apply next.

For every choice of $Y(0) \in (0,\infty)$, almost surely and in L^1 we have that

$$\lim_{t\to\infty}\frac{1}{t}\log X_1(t) \;=\; \varepsilon\lim_{t\to\infty}\frac{1}{t}\int_0^t Y(s)ds \;=\; \varepsilon\int_0^\infty y\,p_\varepsilon(y)\,\mathrm{d}y \;=\; \frac{\varepsilon K_{\alpha-1}\left(4\varepsilon/\sigma^2\right)}{K_\alpha\left(4\varepsilon/\sigma^2\right)}\,, \tag{2.87}$$

where in the first step we have used the first identity in

$$\begin{aligned} X_1(t) \;&=\; X_1(0)\exp\left(\varepsilon\int_0^t Y(s)\,\mathrm{d}s\right)\,,\\ X_2(t) \;&=\; X_2(0)\exp\left(\varepsilon\int_0^t \frac{1}{Y(s)}\,\mathrm{d}s - \alpha\frac{\sigma^2}{2}t + \sigma B_t\right)\,, \end{aligned} \tag{2.88}$$

which is directly derived from (2.65) and holds for all $t > 0$ if both $X_1(0)$ and $X_2(0)$ are positive. The second step in (2.87) is the application of the Pointwise Ergodic Theorem and the last one is an explicit computation. In the same way, by using the second identity in (2.88) we get to (with $x = 4\varepsilon/\sigma^2$)

$$\begin{aligned} \lim_{t\to\infty}\frac{1}{t}\log X_2(t) \;&=\; \varepsilon\lim_{t\to\infty}\frac{1}{t}\left(\int_0^t \frac{1}{Y(s)}ds\right) - \alpha\frac{\sigma^2}{2}\\ &=\; \varepsilon\int_0^\infty \frac{1}{y}\,p_\varepsilon(y)\,\mathrm{d}y - \alpha\frac{\sigma^2}{2}\\ &=\; \frac{\sigma^2}{4}\left(\frac{xK_{1+\alpha}(x)}{K_\alpha(x)} - 2\alpha\right) \overset{(2.75)}{=} \frac{\sigma^2}{4}\frac{xK_{1-\alpha}(x)}{K_\alpha(x)} \;=\; \frac{\varepsilon K_{\alpha-1}\left(4\varepsilon/\sigma^2\right)}{K_\alpha\left(4\varepsilon/\sigma^2\right)}\,, \end{aligned} \tag{2.89}$$

which coincides with what we found in (2.87). This shows that both components have the same exponential growth rate, hence also the norm of $(X_1(t), X_2(t))$, and (2.68) is proven. If instead of starting from the first quadrant, we were starting from the second quadrant, the result is unchanged if we show that the second quadrant is abandoned after a random time that is in $\mathbb{L}^1$. And we have also to show that the diffusion stays in the first quadrant if it starts from there. These facts are somewhat intuitive if we consider the drift

This is somewhat intuitive by considering the drift in (2.83), but it does require some analysis. In this the *Feller test for explosion* turns out to be very useful: we refer to [**7**, pp. 183-184] for the details. □

PROOF OF PROPOSITION 2.11. The proof uses the convergence of the process over bounded times intervals, the ergodic properties of the process for $\Delta > 0$ and some estimates on the invariant measure of the process with $\Delta > 0$ that are uniform in $\Delta \in (0, \Delta_0)$ for some $\Delta_0 > 0$. The proof works under assumptions that are much more general than the ones in Proposition 2.11, see Remark 2.13. For the moment we just refer to [**7**, pp. 211-213] gfor the details of the proof. □

PROOF OF PROPOSITION 2.12. Such a proof is way too heavy and uninteresting, unless you are absolutely fond of Bessel functions. So we refer to [**7**] for the real proof (that heavily exploits several known relations and asymptotic behaviors of special functions [**24**]). At the same time all these results are actually rather elementary, as it will be clear by the following proof of (2.71) in the case of $\alpha \in (0, 1)$ (and without explicit control on the rest):

$$\frac{4}{\sigma^2}\gamma^0_{\sigma,\alpha}(\varepsilon) \overset{x \searrow 0}{\sim} 2\frac{\Gamma(1-\alpha)}{\Gamma(\alpha)}\left(\frac{x}{2}\right)^{2\alpha} . \tag{2.90}$$

For this we first recall that

$$\frac{4}{\sigma^2}\gamma^0_{\sigma,\alpha}(\varepsilon) = \frac{xK_{\alpha-1}(x)}{K_\alpha(x)} , \tag{2.91}$$

and that $K_{\alpha-1}(x) = K_{1-\alpha}(x)$, so (2.90) follows from

$$K_\alpha(x) \overset{x \searrow 0}{\sim} \frac{1}{2}\Gamma(\alpha)\left(\frac{x}{2}\right)^{-\alpha} , \tag{2.92}$$

which holds for $\alpha > 0$. To show (2.92) we simply write directly from (2.66)

$$\begin{aligned} K_\alpha(x) &= \frac{1}{2x^\alpha}\int_0^\infty \frac{1}{y^{1+\alpha}} \exp\left(-\frac{1}{2}\left(x^2 y + \frac{1}{y}\right)\right) \mathrm{d}y \\ &\sim \frac{1}{2x^\alpha}\int_0^\infty \frac{1}{y^{1+\alpha}} \exp\left(-\frac{1}{2y}\right) \mathrm{d}y = \frac{1}{2}\left(\frac{x}{2}\right)^{-\alpha}\int_0^\infty z^{\alpha-1}e^{-z}\,\mathrm{d}z , \end{aligned} \tag{2.93}$$

and the integral in the very last term is $\Gamma(\alpha)$, so (2.92) is proven. □

CHAPTER 3

On Anderson localization in one dimension

3.1. Preliminary facts

Given a sequence of real numbers $(V_n)_{n\in\mathbb{Z}}$ we are going to consider the operator $H = H(V)$ defined by

$$(H\psi)_n \;=\; (H_0\psi)_n + (V\psi)_n \;=\; -\psi_{n+1} - \psi_{n-1} + V_n\psi_n\,, \tag{3.1}$$

for every $\psi \in \mathbb{C}^{\mathbb{Z}}$ and every $n \in \mathbb{Z}$. So H_0 is the discrete Laplacian *without the diagonal term* and V is a *diagonal* operator. It is practical to introduce from now the Wronskian $W(\psi,\phi)$ of $\psi,\phi \in \mathbb{C}^{\mathbb{Z}}$ which is the sequence defined by

$$W_n(\psi,\phi) \;=\; \det\begin{pmatrix}\psi_{n+1} & \phi_{n+1}\\ \psi_n & \phi_n\end{pmatrix} \;=\; \psi_{n+1}\phi_n - \phi_{n+1}\psi_n\,. \tag{3.2}$$

One directly verifies that the following *Green's formula* holds: for every ψ, ϕ and $m \le n$

$$\sum_{k=m}^{n}\left((H\psi)_k\phi_k - \psi_k(H\phi)_k\right) \;=\; W_{m-1}(\psi,\phi) - W_n(\psi,\phi)\,. \tag{3.3}$$

Thanks also to the fact that $\overline{H\psi} = H\overline{\psi}$, Green's formula directly yields that if we choose $D_0 = \{\psi \in \mathbb{C}^{\mathbb{Z}} : |\{n \in \mathbb{Z} : \psi_n \neq 0\}| < \infty\}$ as domain of H, then H is symmetric, that is $\langle\phi, H\psi\rangle = \langle H\phi,\psi\rangle$, where $\langle\phi,\psi\rangle := \sum_n \overline{\phi}_n\psi_n$.

For simplicity and conciseness we are going to choose $\sup_n |V_n| = \|V\|_\infty < \infty$, so it is straightforward to see that H can be extended to $\ell_2(\mathbb{Z}) := \{\psi : \|\psi\|_2^2 = \langle\psi,\psi\rangle < \infty\}$. In fact, H is a bounded self-adjoint operator on $\ell_2(\mathbb{Z})$ and we remark also that D_0 is dense in $\ell_2(\mathbb{Z})$. More quantitatively:

$$\|H\psi\|_2 \;\le\; \|\psi_{\cdot+1}\|_2 + \|\psi_{\cdot-1}\|_2 + \|V_\cdot\psi_\cdot\|_2 \;\le (2 + \|V\|_\infty)\|\psi\|_2\,. \tag{3.4}$$

We are going to look for solutions to $H\psi = E\psi$ for $E \in \mathbb{C}$ and for this the matrix formalism is very useful because one readily sees that if we set for $j \in \mathbb{Z}$

$$Y_j \;:=\; \begin{pmatrix}V_j - E & -1\\ 1 & 0\end{pmatrix}, \tag{3.5}$$

and

$$S_n := \begin{cases} Y_nY_{n-1}\ldots Y_0 & \text{for } n = 0,1,\ldots,\\ Y_n^{-1}Y_{n+1}^{-1}\ldots Y_{-1}^{-1}\ldots & \text{for } n = -1,-2,\ldots,\end{cases} \tag{3.6}$$

then for $n = 0,1,\ldots$

$$\begin{pmatrix}\psi_{n+1}\\ \psi_n\end{pmatrix} \;=\; Y_n\begin{pmatrix}\psi_n\\ \psi_{n-1}\end{pmatrix} = S_n\begin{pmatrix}\psi_0\\ \psi_{-1}\end{pmatrix}, \tag{3.7}$$

and for $n = -1, -2, \dots$

$$\begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix} = S_n \begin{pmatrix} \psi_0 \\ \psi_{-1} \end{pmatrix} . \tag{3.8}$$

Of course, implicit in what we did, was the remark that Y_n is invertible: in fact, $\det(Y_n) = 1$ for every n. We also remark that

$$Y_n^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & V_n - E \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} V_n - E & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.9}$$

that tells us that the Lyapunov exponents associated to (Y_n) and (Y_n^{-1}) coincide.

This matrix representation of the solutions shows that the space of solutions to $H\psi = E\psi$ – all the solutions, not only the ones in $\ell_2(\mathbb{Z})$ – is a two dimensional subspace of $\mathbb{C}^{\mathbb{Z}}$ that can be parametrized by the values of the two solutions on two neighboring sites. In fact, it is practical to choose the two independent solutions $p = p(E)$ and $q = q(E)$ such that

$$p_0(E) = q_{-1}(E) = 1 \ \text{ and } \ q_0(E) = p_{-1}(E) = 0 , \tag{3.10}$$

so

$$\psi_n = p_n(E)\psi_0 + q_n(E)\psi_{-1} . \tag{3.11}$$

REMARK 3.1. *It is also useful to recall that if* $H\psi = E\psi$ $(\psi \not\equiv 0)$ *and* $H\phi = E\phi$, *then* $W_n(\psi, \phi)$ *does not depend on* n *and* $W_n(\psi, \phi) = 0$ *if and only if* $\phi = c\psi$ *for some* $c \in \mathbb{C}$. *The fact that this Wronskian is constant follows because for* $n = 0, 1, \dots$

$$\begin{pmatrix} \psi_{n+1} & \phi_{n+1} \\ \psi_n & \phi_n \end{pmatrix} = S_n \begin{pmatrix} \psi_0 & \phi_0 \\ \psi_{-1} & \phi_{-1} \end{pmatrix} , \tag{3.12}$$

and $\det(S_n) = 1$ *(the argument is identical for negative values of* n*). The second statement is therefore just a well known linear algebra fact.*

REMARK 3.2. *Remark 3.1 implies that the eigenvalues of* H *are simple. We say that* E *is an eigenvalue of* H *if there exists* $\psi \in \ell_2$, $\psi \not\equiv 0$, *such that* $H\psi = E\psi$. *We know a priori that the multiplicity of* E *is either* 1 *or* 2, *so we have to exclude that the multiplicity is* 2 *(but the argument we are giving here does not use the fact that we know that the multiplicity is bounded by* 2*). If* ϕ *is another eigenvalue then* $W_n(\psi, \phi) = C$ *for every* n. *But* $\psi, \phi \in \ell_2$ *readily yields* $|\sum_n W_n(\psi, \phi)| < \infty$, *that is* $C = 0$, *which means that* $\phi \propto \phi$.

3.2. A mathematical approach to the physical viewpoint

Let us first stress that if we were really to talk about quantum mechanics we should study the solutions to the Schroedinger equation (with unit mass and Plank constant equal to one too)

$$i\partial_t \psi(t) = H\psi(t) , \tag{3.13}$$

with suitable initial condition that corresponds to the experiment we intend to perform. We will instead concentrate on the time independent solutions to $H\psi = E\psi$. These two problems are of course related, but we will not discuss the time dependent case (that would be needed to fully justify some of the assertions, from a physical viewpoint).

Moreover, for the physical intuition, it is probably better to work with the notation

$$(H\psi)_n = -\psi_{n+1} - \psi_{n-1} + 2\psi_n + V'_n\psi_n \,, \tag{3.14}$$

where $V'_n = V_n - 2$. Note that for $V' \equiv 0$, then $H\psi = E\psi$ has solutions that do not grow exponentially at $+\infty$ or $-\infty$ if and only if $E \in [0, 4]$: for $E \in (0, 4)$ these solutions are actually bounded and they are the plane waves $\exp(\pm ikn)$, $k \in [0, 2\pi)$, and $E = 2(1 - \cos(k))$.

REMARK 3.3. *We briefly discuss the classical (i.e., non quantum) case: analogous to the problem we consider is the case of a particle of mass one that evolves in a potential $V(x)$, with $V(x) = 0$ for $x \le 0$ and $x \ge L$, and $V(x) > 0$ in $(0, L)$. You should think of something close to a square barrier (because it is closer to what we do next), but we may consider any potential $V(\cdot)$ in C^1 with $\max_x V(x) = U > 0$ and such that $V'(x) < 0$ for $x \in (L-c, L)$ and $V(L-c) = U$. Then $\ddot{x}(t) = -V'(x(t))$, with $x(0) > L$ and $\dot{x}(0) = -v_0 < 0$. By integrating $\dot{x}(t)\ddot{x}(t) = -\dot{x}(t)V'(x(t))$ and by using the fact that the potential is zero at the starting point we obtain $(\dot{x}(t))^2 = v_0^2 - 2V(x(t))$ which tells us that the particle will go past the bump if and only if the kinetic energy of the particle at time zero, i.e. $v_0^2/2$, is larger than the barrier, i.e. U. Otherwise the particle will bounce back or, when $v_0^2/2 = U$, the particle stops at the top of the barrier. This is simply because if t_0 is such that $v_0^2 = 2V(x(t_0)$, then $\ddot{x}(t_0) = -V'(x(0))$ which is strictly positive as long as $v_0^2/2 < U$.*

3.2.1. The free Hamiltonian ($V \equiv 0$). We go ahead by steps and we stress that we do not use the convention of Fig. 3.1 and of (3.14). In particular, the energy E of a wave will be in $[-2, 2]$ and not in $[0, 4]$. With reference to the caption of Fig. 3.1, the tunneling condition $E < U$ becomes $E < U - 2$.

The first step is to consider the case $V \equiv$ constant, which is equivalent to the case $V \equiv 0$. So $H = H_0$. For $k \in [0, 2\pi)$ we define the $\psi_k^{\rightarrow}$ and $\psi_k^{\leftarrow}$ by $\psi_{k,n}^{\rightarrow} = \exp(ikn)$ and $\psi_{k,n}^{\leftarrow} = \exp(-ikn)$. So for $V \equiv 0$ and $E = -2\cos(k)$ we have

$$H_0\psi_k^{\rightarrow} = E\psi_k^{\rightarrow} \quad \text{and} \quad H_0\psi_k^{\leftarrow} = E\psi_k^{\leftarrow}. \tag{3.15}$$

This actually fully solves $H_0\psi = E\psi$ for $E \in (-2, 2)$ because we have found two independent solutions. Since by direct computation one sees that the solutions to $H_0\psi = E\psi$ for $|E| > 2$ have exponential growth at $+\infty$ or $-\infty$, these values of E do not belong to the spectrum of H_0 (this will be explained in detail later on: see Proposition 3.12). We skip the analysis of the cases $E = \pm 2$. Focusing on the solutions for $|E| < 2$ we observe that they are superpositions of the wave $\psi_k^{\rightarrow}$, that *goes to the right*, and of the wave $\psi_k^{\leftarrow}$ that *goes to the left*. In order to get convinced that this is not just a convention one should actually consider the time evolution, i.e. (3.13), of *wave packets*. This is one of the more physical aspects we will no discuss.

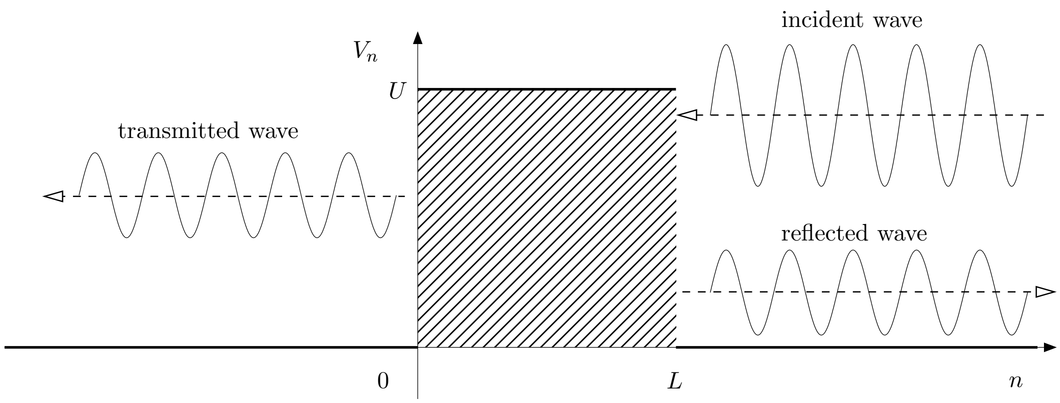

FIGURE 3.1. As we have quickly discussed in Rem. 3.3, when a classical particle encounters an obstacle, it will either cross the obstacle (if its kinetic energy before the barrier is larger than the energy barrier) or it is kicked back (if it is smaller). In quantum mechanics when a particle (or wave) approaches an obstacle (in the image, from the right) there will be a reflected wave and a transmitted wave. If the amplitude of the of the incoming wave of energy E is $\mathtt{i}$ (w.l.o.g., $\mathtt{i} = 1$), the reflected wave is characterized by a complex number $\mathtt{r}$ and the transmitted wave is characterized by another complex number $\mathtt{t}$. We expect that the conservation law $|\mathtt{t}|^2 + |\mathtt{r}|^2 = 1$. Actually (recall the discussion just before Rem. 3.3), waves can have energies $E \in [0, 4]$ (we are using (3.14), not (3.1)!). Here is what can happen:
(1) If $E > U$ then the size $|\mathtt{t}|$ of the transmitted wave does not depend on how large L is: in particular, one can have $|\mathtt{t}| = 1$ or close to 1, hence $|\mathtt{r}| = 0$ or very small, for L arbitrarily large. In fact, we will see that t is never zero.
(2) If $E < U$ then $|\mathtt{t}|$ decays exponentially with L: this is the tunneling effect, i.e. the wave overcomes the obstacle even if its energy is lower than the height the obstacle, but the intensity of the transmitted wave is exponentially small in the length of the obstacle.
If we now introduce randomness in the potential (think of the top of the barrier that, instead of being flat, wiggles) then, no matter how weak the randomness is, we are going to be in the tunneling regime and the transmitted wave has an amplitude which is exponentially weak in L. This is the Anderson localization effect: waves hardly propagate in a disordered potential.

3.2.2. Square potential case: the quantum tunneling effect. We now consider the case is which we introduce an obstacle in our system and we consider the simple case $V_n = U\mathbf{1}_{n\in[1,L]}$, $L \in \mathbb{N}$. This is schematically presented in Fig. 3.1: a wave of wavenumber k, hence energy $E = -2\cos(k)$, comes from $+\infty$ toward the *bump*. What happens is that there will be a reflected wake and a transmitted wave. In order to solve this problem we start from the fact that on the left of the bump there will be only a transmitted wave: $\psi_n = t\exp(-ikn)$ for $n \le 0$, where $t \in \mathbb{C}$. Note that this implies also that $\psi_1 = t\exp(-ik)$. For $n > L$ (and therefore also for $n = L$ as we deduce from (3.1)) we have instead $\psi_n = \exp(-ikn) + r\exp(ikn)$, where $r \in \mathbb{C}$ and $|r|$ is the amplitude of the reflected wave. We just have to solve $H_0\psi_n = (E - U)\psi_n$ for $n = 1, 2, \ldots, L$, with the boundary conditions $\psi_1 = t\exp(-ik)$, $\psi_0 = t$ and $\psi_n = \exp(-ikn) + r\exp(ikn)$ for $n = L$ and $n = L + 1$.

At a physical level, we do expect that the amplitude of the wave that hits the obstacle should match the amplitudes of reflected and transmitted waves, i.e. we expect $|t|^2 + |r|^2 = 1$. We will see that this is true well beyond this first example and it holds also in the fully inhomogeneous setting.

We solve $H\psi_n = (E-U)\psi_n$ by looking for solutions of the form λ^n. It is just a matter of solving

$$\lambda^2 + (E-U)\lambda + 1 = 0 \quad \Longrightarrow \quad \lambda_{1,2} \,=\, \frac{1}{2}\left(U - E \pm \sqrt{(U-E)^2 - 4}\right) . \tag{3.16}$$

Of course these two solutions are also the eigenvalues of Y_n (see (3.5)) when $V_n = U$: since the determinant of this matrix is one we have $\lambda_1\lambda_2 = 1$, which is of course evident also from (3.16). Therefore λ_1 and λ_2 are complex conjugate solutions if $|E-U| < 2$ and in this case they both have absolute value one. So, also *inside the bump*, the solutions are waves, just with a wave number that is not k. On the other hand for $|E-U| > 2$ both solutions are positive: one is larger than one, the other one is smaller than one.

- If $|E-U| < 2$ – i.e. $U' = U-2 < E$, because $E < 2+U$ is trivially verified ($U > 0$) – the solution of $H_0\psi = (E-U)\psi$ is given by $\psi_n = c_\to \exp(ik'n) + c_\leftarrow \exp(-ik'n)$, with $k' \in [0, 2\pi)$ determined by $U - E = 2\cos(k')$ and $n \in \{0, \dots, L+1\}$. Since $\psi_0 = t$ and $\psi_1 = t\exp(-ik)$ we find

$$c_\to \,=\, t\frac{\exp(-ik) - \exp(-ik')}{\exp(ik') - \exp(-ik')} \quad \text{and} \quad c_\leftarrow \,=\, t\frac{\exp(ik') - \exp(-ik)}{\exp(ik') - \exp(-ik')} . \tag{3.17}$$

  In turn, we want also that $\psi_n = \exp(-ikn) + r\exp(ikn)$ for $n = L$ and $n = L+1$. This way we can determine the values of t and r:

$$\begin{aligned} t \,&=\, e^{-iL(k-k')}\frac{\left(e^{2ik}-1\right)\left(e^{2ik'}-1\right)}{1 - e^{2i(k+k'L)} + 2e^{i(k+k'+2k'L)} - 2e^{i(k+k')} + e^{2i(k+k')} - e^{2ik'(L+1)}} , \\ r \,&=\, e^{-i(2kL+k)}\frac{\left(e^{ik}-e^{ik'}\right)\left(e^{i(k+k')}-1\right)\left(e^{2ik'L}-1\right)}{e^{2i(k+k'L)} - 2e^{i(k+2k'L+k')} + 2e^{i(k+k')} - e^{2i(k+k')} + e^{2ik'(L+1)} - 1} , \end{aligned} \tag{3.18}$$

  and one can check that $|t|^2 + |r|^2 = 1$. Note in particular that $|t|$ (and $|r|$) are just periodic functions of L (the period is π/k').
- If $|E-U| > 2$ – i.e. $U' = U - 2 > E$ – the solution of $H\psi = (E-U)\psi$ is given by $\psi_n = c_1\lambda^n + c_2\lambda^{-n}$, with $\lambda > 1 > 0$. Once again we can determine the values of t and r (once again, they satisfy $|t|^2 + |r|^2 = 1$), but in this case

$$\begin{aligned} t \,&=\, \frac{\left(1-e^{2ik}\right)\left(\lambda^2-1\right)e^{-ikL}\lambda^L}{\left(e^{ik}-\lambda\right)^2\lambda^{2L} - \left(e^{ik}\lambda - 1\right)^2} \overset{L\to\infty}{\sim} \lambda^{-L}\frac{\left(1-e^{2ik}\right)\left(\lambda^2-1\right)}{e^{ikL}\left(e^{ik}-\lambda\right)^2} , \\ r \,&=\, -\frac{e^{-2ikL}\left(\lambda^{2L}-1\right)\left(\lambda^2 - 2\lambda\cos(k) + 1\right)}{\left(e^{ik}-\lambda\right)^2\lambda^{2L} - \left(e^{ik}\lambda-1\right)^2} \,=\, \frac{e^{-2ikL+i\pi}}{\left(\left(e^{ik}-\lambda\right)/\left|e^{ik}-\lambda\right|\right)^2} + O\left(1/\lambda^L\right) . \end{aligned} \tag{3.19}$$

  In particular, $1 - |r| = O(1/\lambda^L)$, so the reflection in this case is much stronger and there is a natural *penetration length* ($1/\log\lambda$) of the wave into the obstacle.

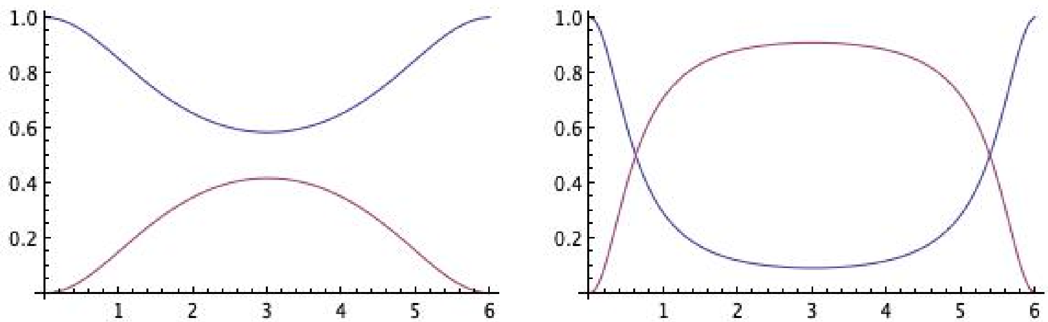

FIGURE 3.2. The plot of $|t|^2$ (in blue) and of $|r|^2$ (in violet) as a function of L over one period. In both cases $k' = \pm\pi/6$, which means $U - E = 2\cos(k') = \sqrt{3}$. The value of k instead is $\pm 2\pi/3$ (i.e., $E = -1$) on the left and $k = \pm\pi/3$ (i.e., $E = 1$) on the right. So in the first case the potential is lower: $U = \sqrt{3} - 1$, with respect to $U = \sqrt{3} + 1$ in the second case.

3.2.3. Disordered potentials. We now consider the case in which V_n is random for $n \in \{1, \ldots, L\}$. More precisely we consider an IID sequence $(U_n)_{n=1,2,\ldots}$ of bounded non trivial, i.e. non constant, random variables, and then we define $V_n := U_n$ for $n = 1, \ldots, L$ and $V_n = 0$ otherwise.

To fix the ideas and as a remarkable example one can think of the case in which U_1 has mean $U > 0$ and very small variance, so we are rather close to the case treated in Section 3.2.2, but we stress that the arguments that follow are general.

We can consider the same set-up of Section 3.2.2. In this case the matrix formalism is going to be very useful. So we consider again an incident wave that comes from $+\infty$ and we foresee the presence of a reflected wave, so $\psi_n = \exp(-ikn) + r\exp(ikn)$ for $k = L, L+1, \ldots$ and, with $E = -2\cos(k)$, $(H\psi)_n = E\psi_n$ for $n = L+1, L+2, \ldots$. Again, to the left of the *obstacle* there will be only the transmitted wave: $\psi_n = t\exp(-ikn)$ for $n = 1, 0, -1, \ldots$ and $(H\psi)_n = E\psi_n$ also for $n = 0, -1, \ldots$ By (3.5)–(3.7) we have for $n = 1, \ldots, L$

$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = Y_n Y_{n-1} \ldots Y_1 \begin{pmatrix} \psi_1 \\ \psi_0 \end{pmatrix}, \tag{3.20}$$

with

$$Y_j := \begin{pmatrix} U_j - E & -1 \\ 1 & 0 \end{pmatrix}, \tag{3.21}$$

and of course $\psi_1 = t\exp(-ik)$ and $\psi_0 = t$. In the next statement we make explicit the dependence on L of t and r by writing $t(L)$ and $r(L)$.

PROPOSITION 3.4. *For every choice of $E \in \mathbb{R}$ the Lyapunov exponent γ of the product of the random matrices (Y_n) is positive. Moreover, if $E \in (-2, 2)$ the transmittal and reflection coefficients $t(L)$ and $r(L)$ are well defined, they satisfy $|t(L)|^2 + |r(L)|^2 = 1$ and, almost surely, $|t(L)| = O(\exp(-\gamma' L))$ for $L \to \infty$ and $\gamma' < \gamma$.*

Therefore if we choose $\mathbb{E}[U_1] > 0$, $|\mathbb{E}[U_1] - E| < 2$ and $\mathrm{var}(U_1)$ very small we are close to the case of the previous section in which *essentially* the wave can go through the obstacle for L arbitrarily large. Now instead, for L large there will be almost no transmitted wave. One can probably come to terms with this by arguing that if the variance is small then $\gamma > 0$ is close to zero, and the penetration length $1/\gamma$ is large. Nevertheless, Proposition 3.4 is telling us that wave propagation is not compatible with the presence of disorder in one dimension.

PROOF OF PROPOSITION 3.4. First of all in the support of the law of U_1 there are a least two values $u_0 \neq u_1$. The corresponding matrices have have the form

$$M_j = \begin{pmatrix} a_j & -1 \\ 1 & 0 \end{pmatrix}, \tag{3.22}$$

with $a_j = E - u_j$. Since

$$M := M_0 M_1^{-1} = \begin{pmatrix} a_0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & a_1 \end{pmatrix} = \begin{pmatrix} 1 & a_0 - a_1 \\ 0 & 1 \end{pmatrix}, \tag{3.23}$$

by taking powers of M we readily see that the group that contains the support of the law of Y_1 is non compact. The irreducibility follows by observing that $\lim_n M^n x / \|M^n x\| = (1, 0)$ and of course the ray associated to this direction is invariant. So, if there is a finite union of rays that is invariant under the action of all the elements of the group spanned by the support of the measure, it must coincide with the ray that goes through $y = (1, 0)$. But this is impossible because $M_j^{-1} y$ is not collinear with y.

Therefore, by Theorem 1.8 (Furstenberg Theorem), we have that $\gamma > 0$.

Now we recall that we are interested in

$$Y_n Y_{n-1} \ldots Y_1 \begin{pmatrix} \cos(k) - i\sin(k) \\ 1 \end{pmatrix}, \tag{3.24}$$

times the transmittal coefficient $t \in \mathbb{C}$. But, again by Theorem 1.8, we know that almost surely the logarithm of the norm of the real part of the vector in (3.24) behaves like $n\gamma$ for n large. In fact, the same is true also also for the imaginary part because we are excluding $k = 0$ and $k = \pi$ (i.e. $E = \pm 2$).

Next step is writing

$$t(L) Y_L Y_{L-1} \ldots Y_1 \begin{pmatrix} \cos(k) - i\sin(k) \\ 1 \end{pmatrix} = \begin{pmatrix} \exp(-ik(L+1)) + r(L)\exp(ik(L+1)) \\ \exp(-ikL) + r(L)\exp(ikL) \end{pmatrix}. \tag{3.25}$$

We claim that also in this case we have $|r(L)|^2 + |t(L)|^2 = 1$ (see Lemma 3.5). Therefore the norm (in $\mathbb{C}^2$: $\|(z, w)\|^2 = |z|^2 + |w|^2$) of the right hand side of (3.25) is bounded by $\sqrt{2}(1 + |r(L)|)$ because the absolute value of each one the two entries of the vector is bounded by $1 + |r(L)|$, while the norm of the left-hand side is equal to

$$|t(L)| \left\| Y_L Y_{L-1} \ldots Y_1 \begin{pmatrix} \cos(k) - i\sin(k) \\ 1 \end{pmatrix} \right\|, \tag{3.26}$$

and we know that the norm that multiplies $|t(L)|$ grows with the correct exponential rate because real and imaginary parts do (for $v \in \mathbb{C}^2$ we have $\|v\|^2 = \|\Re v\|^2 + \|\Im v\|^2$). This implies the claimed (almost sure) exponential decay of $|t(L)|$. $\square$

LEMMA 3.5. *For every $k \in (0, 2\pi) \setminus \{\pi\}$, $L = 0, 1, 2, \dots$ and every choice of the real numbers $a_1, a_2, \dots$ we have that there exists a unique choice of $(t, r) \in \mathbb{C}^2$ such that*

$$t \begin{pmatrix} a_L & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{L-1} & -1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \exp(-ik) \\ 1 \end{pmatrix} = \begin{pmatrix} \exp(-ik(L+1)) + r\exp(ik(L+1)) \\ \exp(-ikL) + r\exp(ikL) \end{pmatrix}, \quad (3.27)$$

where the matrix product should be read as the identity matrix if $L = 0$. Moreover $t \neq 0$ and $|t|^2 + |r|^2 = 1$.

PROOF. The solution exists and it is unique because (3.27) can be written, with obvious definition of the matrix T_L (of determinant 1), as

$$t\, T_L \begin{pmatrix} \exp(-ik) \\ 1 \end{pmatrix} = \begin{pmatrix} \exp(-ik(L+1)) & \exp(ik(L+1)) \\ \exp(-ikL) & \exp(ikL) \end{pmatrix} \begin{pmatrix} 1 \\ r \end{pmatrix} =: A_{L+1} \begin{pmatrix} 1 \\ r \end{pmatrix}, \quad (3.28)$$

and the matrix A_{L+1} has determinant $e^{-ik} - e^{ik} = -2i\sin(k) \neq 0$. Therefore with $B := A_{L+1}^{-1} T_L$

$$t\, B \begin{pmatrix} \exp(-ik) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ r \end{pmatrix}, \quad (3.29)$$

Since $\det(B) \neq 0$ we see that $t \neq 0$. We therefore have an expression for the vector $(1/t, r/t)$ and r and t are uniquely identified.

We are now at the key point of showing that $|t|^2 + |r|^2 = 1$. We do this by looking at the action of one matrix at time, see Figure 3.3. For $\ell = 1, \dots, L$, but in fact ℓ can be seen as an arbitrary integer number, the action of the matrix on the wave function between $\ell - 1$ and ℓ on the basis of $\psi^{\leftarrow}_{k,\cdot}$ and $\psi^{\rightarrow}_{k,\cdot}$ yields

$$\begin{pmatrix} v & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a'e^{-ik\ell} + b'e^{ik\ell} \\ a'e^{-ik(\ell-1)} + b'e^{ik(\ell-1)} \end{pmatrix} = \begin{pmatrix} ae^{-ik(\ell+1)} + be^{ik(\ell+1)} \\ ae^{-ik\ell} + be^{ik\ell} \end{pmatrix} \quad (3.30)$$

where where $v = U_{\ell+1} - E$ is at this stage just and arbitrary real number. Note that the (complex) coefficients a and a' are the amplitudes of the waves going from right to left, respectively before and after going through the potential *slice* at ℓ. In the specular way, b and b' are the amplitudes of the waves going from left to right, respectively after and before going through the potential. What we want to show is that, regardless of the value of v, we have the conservation law $|a|^2 - |b|^2 = |a'|^2 - |b'|^2$ which is possibly more clear in the non symmetric version $|a|^2 + |b'|^2 = |a'|^2 + |b|^2$ which means that what goes in must come out. We rewrite (3.30) as

$$\begin{pmatrix} a \\ b \end{pmatrix} = A_{\ell+1}^{-1} \begin{pmatrix} v & -1 \\ 1 & 0 \end{pmatrix} A_\ell \begin{pmatrix} a' \\ b' \end{pmatrix} \quad (3.31)$$

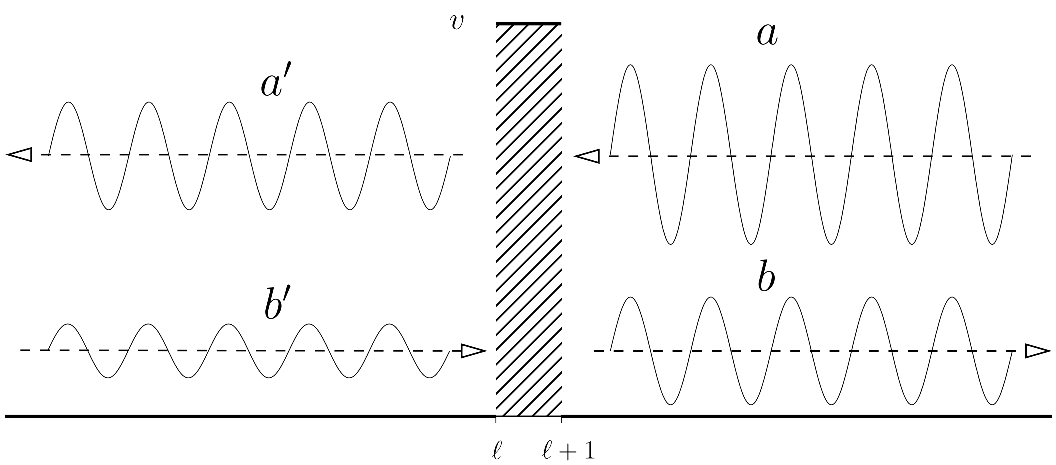

FIGURE 3.3. The effect of a slice of potential. The waves going from right to left are two: one, with intensity a, is incoming and one, with intensity a' is outgoing (transmitted). In the same way the waves from left to right are an incoming one with intensity b' and and outgoing one with intensity b. Therefore we expect $|a|^2 + |b'|^2 = |a'|^2 + |b|^2$.

where as in (3.28)

$$A_\ell = A_\ell(k) = \begin{pmatrix} e^{-ik\ell} & e^{ik\ell} \\ e^{-ik(\ell-1)} & e^{ik(\ell-1)} \end{pmatrix}, \tag{3.32}$$

and we recall that $\det(A_\ell) = -2i\sin(k)$ (so $\det(A_\ell^*) = 2i\sin(k)$). Therefore $|a|^2 - |b|^2 = |a'|^2 - |b'|^2$ holds if

$$A_\ell^* \begin{pmatrix} v & 1 \\ -1 & 0 \end{pmatrix} \left(A_{\ell+1}^{-1}\right)^* \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A_{\ell+1}^{-1} \begin{pmatrix} v & -1 \\ 1 & 0 \end{pmatrix} A_\ell = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} =: D. \tag{3.33}$$

It is therefore a matter of a computation (remark that the result does not depend on the value of ℓ):

$$\begin{aligned} \left(A_\ell^{-1}\right)^* D A_\ell^{-1} &= \frac{1}{4(\sin(k))^2} \begin{pmatrix} e^{-ik(\ell-1)} & -e^{ik(\ell-1)} \\ -e^{-ik\ell} & e^{ik\ell}, \end{pmatrix} \begin{pmatrix} e^{ik(\ell-1)} & -e^{ik\ell} \\ e^{-ik(\ell-1)} & -e^{-ik\ell} \end{pmatrix} \\ &= \frac{1}{4(\sin(k))^2} \begin{pmatrix} 0 & -2i\sin(k) \\ 2i\sin(k) & 0 \end{pmatrix} \\ &= \frac{1}{2i\sin(k)} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \end{aligned} \tag{3.34}$$

so

$$\begin{aligned} \begin{pmatrix} v & 1 \\ -1 & 0 \end{pmatrix} \left(A_{\ell+1}^{-1}\right)^* D A_{\ell+1}^{-1} \begin{pmatrix} v & -1 \\ 1 & 0 \end{pmatrix} &= \frac{1}{2i\sin(k)} \begin{pmatrix} v & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} v & -1 \\ 1 & 0 \end{pmatrix} \\ &= \frac{1}{2i\sin(k)} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \end{aligned} \tag{3.35}$$

Finally

$$
\begin{aligned}
A_\ell^* \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} A_\ell &= \begin{pmatrix} e^{ik\ell} & e^{ik(\ell-1)} \\ e^{-ik\ell} & e^{-ik(\ell-1)} \end{pmatrix} \begin{pmatrix} e^{-ik(\ell-1)} & e^{ik(\ell-1)} \\ -e^{-ik\ell} & -e^{ik\ell} \end{pmatrix} \\
&= 2i\sin(k) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\end{aligned} \tag{3.36}
$$

Therefore (3.33) holds and we are done. □

3.3. Operator viewpoint on localization

Let us first introduce some important tools in our analysis. First of all, the spectrum $\Sigma(A)$ of a bounded operator A is the complement of the set of $\lambda \in \mathbb{C}$ such that $A - \lambda I$ (I is the identity) has an inverse: note that, since A is bounded, such an inverse is necessarily bounded. $\Sigma(A)$ is closed, bounded and not empty.

We are focusing on $A = H$ and H self-adjoint (and we recall that we work on $\ell_2 = \ell_2(\mathbb{Z})$). Therefore $\Sigma(H) \subset \mathbb{R}$. If H is random, a priori also $\Sigma(H)$ is random. However:

PROPOSITION 3.6. *If (V_n) is an IID sequence then for almost all realizations of this sequence we have that $\Sigma(H) = [-2, 2] + \mathrm{Supp}(P_V) := \{x + y : x \in [-2, 2],\ y \in \mathrm{Supp}(P_V)\}$, where P_V is the law of V_1.*

PROOF. Recall the definion of $H = H_0 + V$ in (3.1). Let us first point out that $[-2, 2]$ is the spectrum of the operator H_0 as can be seen by direct computation, plus some operator theory considerations. In fact it suffices to remark that $\|H_0\| = \sup_{\psi:\,\|\psi\|=1} \|H_0\psi\| \le 2$ so $\Sigma(H_0) \subset [-2, 2]$ and, for $\psi_n = \exp(ikn)$, $H_0\psi = -2\cos(k)\psi$, so $\Sigma(H_0) \supset [-2, 2]$. Note that $\psi \notin \ell_2$, but if we define $\psi^{(L)}$ by setting $(\psi^{(L)})_n = \psi_n/\sqrt{2L+1}$ if $|n| \le L$ and $(\psi^{(L)})_n = 0$ otherwise, then one readily sees that $\|H_0\psi^{(L)} + 2\cos(k)\psi^{(L)}\| = O(1/\sqrt{L})$ which means that $-2\cos(k) \in \Sigma(H_0)$. Not only, this actually means that $-2\cos(k)$ is not an isolated eigenvalue of finite multiplicity (more on the spectrum of H_0 in Remark 3.8). All of this is consequence of the following fundamental result that we will repeatedly use (we give it in the restricted framework of self-adjoint operators, see for example [**28**, Lemma 2.16 and Lemma 6.17] for a proof and more):

LEMMA 3.7 (Weyl's criterion). *For T self-adjoint, $\lambda \in \Sigma(T)$ if and only if there exists a sequence of $(\psi^{(n)})$, called "Weyl sequence", with $\|\psi^{(n)}\| = 1$ for every n such that*

$$
\lim_n \left\|(T - \lambda I)\psi^{(n)}\right\| = 0\,. \tag{3.37}
$$

Moreover, $\lambda \in \Sigma(T)$ is not an isolated eigenvalue of finite multiplicity if and only if there exists $(\psi^{(n)})$ with $\|\psi^{(n)}\| = 1$ such that (3.37) holds and $\lim_n(\varphi, \psi^{(n)}) = 0$ for every φ such that $\|\varphi\| < \infty$. In this case $(\psi^{(n)})$ is called "singular Weyl sequence".

The important part of the statement for us is the first one. The second part is included as a *sanity check*: we stress that by isolated eigenvalue we mean isolated in the whole spectrum, i.e. that there exists $\varepsilon > 0$ such that $\Sigma(T) \cap \{z : |z - \lambda| < \varepsilon\} = \{\lambda\}$. But in reality we will prove Proposition 3.6, which implies that there is no isolated point in $\Sigma(H)$ (otherwise said, no subset of the spectrum is discrete, or that there is no discrete spectrum), without using the second part of Lemma 3.7.

The operator V instead is diagonal in the orthonormal basis $(e^k)_{k\in\mathbb{Z}}$, defined by $e^k_n = \delta_{n,k}$ (Kronecker delta). If is then clear that, almost surely, the spectrum of V is $\mathrm{Supp}(P_V)$. Note that the spectrum of V is never discrete, not even if the sequence (V_k) takes only a finite number of values. This is due to the fact that there are either eigenvectors of infinite multiplicity and/or the eigenvalues are not isolated. This can be directly seen, but it is helpful to see it also through Lemma 3.7: for example, if (V_k) takes only a finite number of values, the some value λ is taken up infinitely many times, i.e. $V_k = \lambda$ for $k \in I = \{i_1, i_2, \ldots\} \subset \mathbb{Z}$, $|I| = \infty$. So the $\sum_{k=1}^n e^{i_k}/\sqrt{n}$ forms a singular Weyl sequence (of eigenfunctions this time!).

Both H_0 and V are bounded self-adjoint operators, as $H = H_0 + V$ is. In this context it is not difficult to see that $\max \Sigma(H) \le \max \Sigma(H_0) + \max \Sigma(V)$ and that $\min \Sigma(H) \ge \min \Sigma(H_0) + \min \Sigma(V)$. It is in fact a direct consequence of the following statement for self-adjoint operators (see [**28**, Th. 2.20])

$$\max \Sigma(T) = \sup_{\psi:\, \|\psi\|=1} (\psi, T\psi) \qquad \text{and} \qquad \min \Sigma(T) = \inf_{\psi:\, \|\psi\|=1} (\psi, T\psi)\,. \tag{3.38}$$

This is sufficient to show that $\Sigma(H) \subset [-2, 2] + \mathrm{Supp}(P_V)$ only when this set is an interval. But $\Sigma(H) \subset [-2, 2]+\mathrm{Supp}(P_V)$ does hold in general: it follows for example by using [**19**, Ch. V, Th. 4.10].

The opposite inclusion follows from the first part of Weyl criterion (Lemma 3.7)). In fact it is straightforward to see that almost surely for every $E \in \mathrm{Supp}(P_V)$, every $L \in \mathbb{N}$ and every $\varepsilon > 0$ one can find $N \in \mathbb{Z}$ such that $V_n \in [E - \varepsilon, E + \varepsilon]$ for every $n = N+1, N+2, \ldots, N+L$. We then consider, for $k \in [0, 2\pi)$, $\psi_n = \exp(ikn)/\sqrt{L})$ for $n = N + 1, N + 2, \ldots, N + L$, and $\psi_n = 0$ otherwise. It is therefore easy to see that

$$\|(H - EI)\psi\| \le \varepsilon + O(1/\sqrt{L})\,, \tag{3.39}$$

where the $O(1/\sqrt{L})$ correction comes from the contribution near 0 and L. Therefore, by choosing a sequence of (L_n) going to ∞ and of (ε_n) that goes to zero we see that $E-2\cos(k)$ belongs to $\Sigma(H)$. This completes the proof, but we add the observation that of course the Weyl sequence we built is singular, since there is no isolated point in the spectrum. □

Now we keep going for a while with (V_n) an arbitrary bounded real sequence. Since H is a bounded operator, $f(H)$ is directly defined for every polynomial function $f(\cdot)$ and, by approximation, the definition can be extended to $f \in C^0(\mathbb{R}; \mathbb{R})$. Therefore, thanks to the Riesz-Markov Theorem, for every $\psi \in \ell_2$ there exists a unique measure μ_ψ – we call it *spectral measure* associated to ψ – such that

$$(f(H)\psi, \psi) = \int_{\mathbb{R}} f(\lambda)\mu_\psi(\,\mathrm{d}\lambda)\,, \tag{3.40}$$

for every $f \in C^0(\mathbb{R};\mathbb{R})$. Let us remark that the support of μ_ψ is contained in $\Sigma(H)$: in fact, since the spectrum is a closed set, if $\lambda_0 \in \mathbb{R}\setminus\Sigma(H)$ then $\lambda \in \mathbb{R}\setminus\Sigma(H)$ if $|\lambda-\lambda_0| \le \delta$ for some $\delta > 0$. Hence $C := \sup_{\lambda:\,|\lambda-\lambda_0|\le\delta} \|(H\psi-\lambda\psi)^{-1}\| < \infty$ so, by applying a cut-off to $x \mapsto 1/x^2$, one sees that $\int_{\mathbb{R}}(\lambda'-\lambda)^{-2}\mu_\psi(\mathrm{d}\lambda') \le C^2$ which implies $\int_{\lambda_0-\varepsilon}^{\lambda_0+\varepsilon}\int_{\mathbb{R}}(\lambda'-\lambda)^{-2}\mu_\psi(\mathrm{d}\lambda')\,\mathrm{d}\lambda \le 2\varepsilon C^2$. But $\int_{\lambda_0-\varepsilon}^{\lambda_0+\varepsilon}(\lambda'-\lambda)^{-2}\,\mathrm{d}\lambda = \infty$ for every $\lambda' \in (\lambda_0-\varepsilon, \lambda_0+\varepsilon)$ and this implies that $\mu_\psi((\lambda_0-\varepsilon,\lambda_0+\varepsilon)) = 0$.

In fact, $f(H)$ is defined also for less regular functions, notably indicator functions of measurable sets and we have

$$\mu_\psi(B) \,=\, (\mathbf{1}_B(H)\psi, \psi)\,, \tag{3.41}$$

for every Borel subset B of $\mathbb{R}$. Note that μ_ψ is a positive measure also if ψ takes complex values.

In the same way for every $\psi, \varphi \in \ell_2$ we can introduce the measure $\mu_{\psi,\varphi}$ defined by

$$\mu_{\psi,\varphi}(B) \,=\, (\mathbf{1}_B(H)\psi, \varphi)\,, \tag{3.42}$$

which is a complex measure if ψ and φ take values in $\mathbb{C}$.

Another important result we borrow from functional analysis is that the Hilbert space ℓ_2 can be decomposed into three orthogonal subspaces that are H invariant: with $\ll$ to denote *absolutely continuous with respect to* and λ the Lebesgue measure on $\mathbb{R}$ we introduce

$$\begin{aligned} \hbar^{\mathrm{a}} &:= \{\psi \in \ell_2 : \mu_\psi \ll \lambda\}\,, \quad \hbar^{\mathrm{s}} := \{\psi \in \ell_2 : \mu_\psi \not\ll \lambda \text{ and } \mu_\psi(\{x\}) = 0\,\forall x\}\,, \\ \hbar^{\mathrm{p}} &:= \left\{\psi \in \ell_2 : \mu_\psi = \sum_j c_j\delta_{x_j} \text{ for a suitable choice of } (c_j) \text{ and } (x_j)\right\}. \end{aligned} \tag{3.43}$$

Note that here we have exploited the *Lebesque decomposition* of a measure:

$$\mu_\psi \,=\, \mu_\psi^{\mathrm{a}} + \mu_\psi^{\mathrm{s}} + \mu_\psi^{\mathrm{p}}\,, \tag{3.44}$$

i.e., that every measure on $\mathbb{R}$ can be written as sum of three mutually singular measures: one absolutely continuous with respect to the Lebesgue measure, one singular with respect to Lebesgue (but for which the measure of every point is zero) and one that is *pure point.* We can therefore consider the spectrum of the restriction of H to each one these three subspaces and

$$\Sigma(H) \,=\, \Sigma^{\mathrm{a}}(H) \cup \Sigma^{\mathrm{s}}(H) \cup \Sigma^{\mathrm{p}}(H)\,. \tag{3.45}$$

In general these three spectra – absolutely continuous, singular, pure point – are not disjoint.

REMARK 3.8. *Let us discuss the spectrum of H_0 as an example. The spectrum of H_0 can be understood in detail thanks to the fact that H_0 (on ℓ_2) is unitarily equivalent to the operator $\widehat{H}_0$ on $\mathbb{L}^2$ of $[0,2\pi)$ equipped with the Lebesgue measure defined by $\widehat{H}_0 f(k) = -2\cos(k)f(k)$. The unitary operator that links the two spaces is of course the Fourier Transform $\widehat{\psi}(k) = \sum_n \exp(ikn)\psi_n/\sqrt{2\pi}$. By exploiting this equivalence it becomes evident that $[-2,2]^{\complement}$ is not in the spectrum: to show that $\Sigma(H_0) = [-2,2]$ requires in any case Weyl's criterion (Lemma 3.7). But we can* *also make the spectral measure associated to ψ explicit. In fact for $\psi \in \ell_2$ and $t \in [-2,2] = \Sigma(H_0)$*

$$\mu_\psi((-\infty,t]) \,=\, \big(\mathbf{1}_{[-2,t]}(H_0)\psi,\psi\big) \,=\, \Big(\mathbf{1}_{[-2,t]}\left(\widehat{H_0}\right)\widehat{\psi},\widehat{\psi}\Big)$$

$$= \int_0^{2\pi} |\widehat{\psi}(k)|^2 \mathbf{1}_{-2\cos(k)\le t}\,\mathrm{d}k \,=\, \int_0^{\arccos(-t/2)} |\widehat{\psi}(k)|^2\,\mathrm{d}k + \int_{2\pi-\arccos(-t/2)}^{2\pi} |\widehat{\psi}(k)|^2\,\mathrm{d}k\,, \tag{3.46}$$

so μ is absolutely continuous with respect to the Lebesgue measure with density

$$\mathbf{1}_{(-2,2)}(t)\left(\left|\widehat{\psi}(\arccos(-t/2))\right|^2 + \left|\widehat{\psi}(2\pi-\arccos(-t/2))\right|^2\right)\Big/\left(2\sqrt{1-t^2/4}\right)\,. \tag{3.47}$$

Hence the spectrum of H_0 is absolutely continuous: $\Sigma(H_0) = \Sigma^{\mathrm{a}}(H_0)$ and $\Sigma^{\mathrm{s}}(H_0) = \Sigma^{\mathrm{p}}(H_0) = \emptyset$.

The definition (3.45) depends a priori on the choice of ψ. But it is possible to make a *canonical* (and, in a sense, also *optimal*) choice for the spectral measure. For this we recall the orthonormal basis $(e^k)_{k\in\mathbb{Z}}$, defined by $e^k_n = \delta_{n,k}$ used above. Note that $\big(\psi, e^k\big) = \psi_k$. Moreover we use the notations $\mu_{n,k}$ for μ_{e^n,e^k} and μ_n for μ_{e^n}. Here is our choice for the spectral measure:

$$\mu \,=\, \mu_{-1} + \mu_0\,. \tag{3.48}$$

Of course μ is not a probability, but, since $\mu/2$, is we will at times write $\mu(\mathrm{d}\lambda)$-almost surely, referring to $\mu/2$. Recall now the definition of $p_n(\lambda)$ and $q_n(\lambda)$ from (3.10). In particular, $Hp(\lambda) = \lambda p(\lambda)$ and $Hq(\lambda) = \lambda q(\lambda)$.

LEMMA 3.9. *$p_n(H)e^0 + q_n(H)e^{-1} = e^n$ for every $n \in \mathbb{Z}$.*

PROOF. $p_0(H) = q_{-1}(H)$ is the identity operator and $q_0(H) = p_{-1}(H)$ is multiplication by 0. So the cases $n = 0$ and $n = -1$ hold. We then proceed by induction in the two directions by exploiting that for every k we have

$$He^k \,=\, -e^{k+1} - e^{k-1} + V_k e^k\,. \tag{3.49}$$

So we want to show $p_m(H)e^0 + q_m(H)e^{-1} = e^m$ for $m = n+1$ if we know it for $m = n$ and $m = n-1$, with $n = 0,1,2,\ldots$ (respectively, for $m = n-1$ if we know it for $m = n$ and $m = n+1$, with $n = -1,-2,\ldots$). This is just a matter of applying H to $p_m(H)e^0 + q_m(H)e^{-1} = e^m$ and of using repeatedly (3.49). Here is one of the steps (*going down*): we know the result for n and $n+1$ so $Hp_n(H)e^0 + Hq_n(H)e^{-1} = He^n$ can be developed and reordered into

$$p_{n-1}(H)e^0 + q_{n-1}(H)e^0 - e^{n-1} \,=$$
$$-\,p_{n+1}(H)e^0 - q_{n+1}(H)e^0 + e^{n+1} + V_n\Big(p_n(H)e^0 + q_n(H)e^0 - e^n\Big)\,. \tag{3.50}$$

Since the right-hand side is zero by the induction assumption, the proof is complete. □

Here is why μ may be considered the spectral measure of interest for our problem.

LEMMA 3.10. *For every $\psi \in \ell_2$ we have that $\mu_\psi \ll \mu$.*

PROOF. First of all $0 = \mu(B) = \mu_{-1}(B) + \mu_0(B)$ implies that

$$\|\mathbf{1}_B(H)e^0\|^2 \,=\, \big(\mathbf{1}_B(H)e^0, \mathbf{1}_B(H)e^0\big) \,=\, \big(\mathbf{1}_B(H)e^0, e^0\big) \,=\, \mu_0(B) \,=\, 0\,, \tag{3.51}$$

so $\mathbf{1}_B(H)e^0 = 0 \in \ell_2$ and the very same argument yields also $\mathbf{1}_B(H)e^{-1} = 0$. But then Lemma 3.9 directly yields that $\mathbf{1}_B(H)e^n = 0$ for every n. Since (e^n) is a basis of ℓ_2 we directly have the claim (in fact, $\mathbf{1}_B(H) = 0$). □

PROPOSITION 3.11. *We have that $\Sigma^{\mathrm{a}}(H) = \mathrm{Supp}(\mu^{\mathrm{a}})$, $\Sigma^{\mathrm{p}}(H) = \mathrm{Supp}(\mu^{\mathrm{p}})$ and $\Sigma^{\mathrm{s}}(H) = \mathrm{Supp}(\mu^{\mathrm{s}})$. Moreover E is an eigenvalue of H (i.e., there exists $\psi \in \ell_2$ non zero such that $H\psi = E\psi$) if and only if $\mu(\{E\}) > 0$.*

PROOF OF PROPOSITION 3.11. This is based on the first part of Weyl's criterion (Lemma 3.7).First of all we remark that for the three spectrum-support equalities it suffices to give the argument for $\Sigma(H) = \mathrm{Supp}(\mu)$: the same argument can be repeated for the three different orthogonal spaces (that are H-invariant). So $E \in \Sigma(H)$ if and only if $\lim_n \int (t-E)^2 \mu_{\psi_n}(\mathrm{d}t) = 0$ with $\psi_n \in H$ and $\|\psi_n\| = 1$, i.e. $\mu_{\psi_n}(\mathbb{R}) = 1$. In particular we have that $\mu_{\psi_n} \Rightarrow \delta_E$. Now if $E \in \Sigma(H)$ then for every $\varepsilon > 0$ we have that $\mu_{\psi_n}((E-\varepsilon, E+\varepsilon)) \ge 1/2$ for n sufficiently large, which implies that $\mu((E-\varepsilon, E+\varepsilon)) > 0$ by Lemma 3.10, so $E \in$Supp(μ). On the other hand, if $E \not\in$ Supp(μ) then $\mu((E-\varepsilon, E+\varepsilon)) = 0$ for some $\varepsilon > 0$, hence, always by Lemma 3.10, $\mu_{\psi_n}((E-\varepsilon, E+\varepsilon)) = 0$ for every n and therefore $\mu_{\psi_n} \not\Rightarrow \delta_E$, hence $E \not\in \Sigma(H)$.

We are left with showing that E is an eigenvalue of H if and only if $\mu(\{E\}) > 0$. We know that, for every $\psi \in \ell_2$, $\|(H - EI)\psi\| = \int (t-E)^2 \mu_\psi(\mathrm{d}t)$. So ψ is an eigenfunction with eigenvalue E if an only if $\mu_\psi(\{E\}^{\complement}) = 0$ or, equivalently, $\mu_\psi \propto \delta_E$. Therefore, by Lemma 3.10, ψ is an eigenfunction with eigenvalue E if an only if $\mu(\{E\}) > 0$. □

It is now practical to introduce a *measure valued matrix* associated to H:

$$\mathtt{M} := \begin{pmatrix} \mu_0 & \mu_{-1,0} \\ \mu_{0,-1} & \mu_{-1} \end{pmatrix} . \tag{3.52}$$

We note that this matrix is symmetric and that its trace is μ. A less immediate observation is that

$$\mu_{m,n}(\mathrm{d}E) \,=\, (p_m(E), q_m(E))\mathtt{M}(\mathrm{d}E) \begin{pmatrix} p_n(E) \\ q_n(E) \end{pmatrix} , \tag{3.53}$$

which follows because for every Borelian B by Lemma 3.9 we have

$$
\begin{aligned}
\mu_{m,n}(B) = (\mathbf{1}_B(H)e^m, e^n) \;&=\; (\mathbf{1}_B(H)e^m, \mathbf{1}_B(H)e^n) \\
&= \big(p_m(H)\mathbf{1}_B(H)e^0 + q_m(H)\mathbf{1}_B(H)e^{-1}, p_n(H)\mathbf{1}_B(H)e^0 + q_n(H)\mathbf{1}_B(H)e^{-1}\big) \\
&= \int_B p_m(E)p_n(E)\mu_0(\,\mathrm{d}E) + \int_B q_m(E)q_n(E)\mu_{-1}(\,\mathrm{d}E) \\
&\qquad + \int_B (p_m(E)q_n(E) + p_n(E)q_m(E))\mu_{0,-1}(\,\mathrm{d}E)\,.
\end{aligned}
\tag{3.54}
$$

It is actually practical to use that $\mathtt{M} \ll \mu$ and work with the density $M(\cdot)$ of $\mathtt{M}$ with respect to μ. Without surprise, $S(E)$ is symmetric $\mu(\,\mathrm{d}E)$-almost everywhere. Moreover it is also non negative (as a matrix), which is a direct consequence of the fact that $\mathtt{M}(B) \geq 0$ as a matrix, for every Borelian B, and this last fact follows by observing that for every $x \in \mathbb{R}^2$

$$
\begin{aligned}
\langle x, \mathtt{M}(B)x\rangle \;&=\; x_1^2\mu_0(B) + x_1x_2\left(\mu_{0,-1}(B) + \mu_{-1,0}(B)\right) + x_2^2\mu_{-1}(B) \\
&=\; \|x_1\mathbf{1}_B(H)\delta^0 + x_2\mathbf{1}_B(H)\delta^{-1}\|^2 \;\geq\; 0\,.
\end{aligned}
\tag{3.55}
$$

We are now ready to prove an important result about *generalized eigenvectors.*

PROPOSITION 3.12. *$\mu(\,\mathrm{d}E)$-a.s. there exists ψ such that $H\psi = E\psi$ and, for every $\varepsilon > 0$, $\psi_n = O(n^{\frac{1}{2}+\varepsilon})$ for $|n| \to \infty$.*

PROOF. In the whole proof we work $\mu(\,\mathrm{d}E)$-almost surely. As we have seen, $M(E)$ is symmetric and non negative. Moreover, since the trace of $\mathtt{M}$ is the measure μ, the trace of $M(E)$ is one. So there exists a (real) orthogonal matrix $A = A(E)$ such that

$$
A\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} A^*\,,
\tag{3.56}
$$

with $a \geq b \geq 0$ and $a + b = 1$, so $a \geq 1/2$. Therefore if we use (3.53) for $n = m$ and hiding the dependence on E we obtain

$$
\mathrm{d}\mu_n \;=\; (p_n, q_n)A\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} A^*\begin{pmatrix} p_n \\ q_n \end{pmatrix}\,\mathrm{d}\mu\,.
\tag{3.57}
$$

The key point now is to remark that the two entries of the vector $(p_n, q_n)A$, for $n \in \mathbb{Z}$, are the solutions to $H\psi = E\psi$ with *conditions at the origin* given by the columns of A (recall (3.11)): this becomes clearer if we write write

$$
A \;=\; \begin{pmatrix} \psi_0 & \varphi_0 \\ \psi_{-1} & \varphi_{-1} \end{pmatrix}\,,
\tag{3.58}
$$

so $(\psi_n, \varphi_n) = (p_n, q_n)A$, and $\psi_n = \psi_n(E)$, respectively $\varphi_n = \varphi_n(E)$, is the solution of $H\psi = E\psi$ with ψ_0 and ψ_{-1} given, respectively solution of $H\varphi = E\varphi$ with φ_0 and

φ_{-1} given. Therefore we readily see that

$$(p_n, q_n) A \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} A^* \begin{pmatrix} p_n \\ q_n \end{pmatrix} = (\psi_n, \varphi_n) \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} \psi_n \\ \varphi_n \end{pmatrix} \geq \frac{1}{2} \psi_n^2 . \tag{3.59}$$

On the other hand $\mu_n(\mathbb{R}) = (e^n, e^n) = 1$ so $\frac{1}{2} \int \psi_n^2 \, \mathrm{d}\mu \leq 1$ and therefore by the Fubini-Tonelli Theorem for every $\varepsilon > 0$ we have

$$\int_{\mathbb{R}} \sum_{n \neq 0} \frac{(\psi_n(E))^2}{|n|^{1+\varepsilon}} \mu(\,\mathrm{d}E) \leq 2 \sum_{n \neq 0} \frac{1}{|n|^{1+\varepsilon}} < \infty . \tag{3.60}$$

We can therefore conclude that $\mu(\,\mathrm{d}E)$-a.s.

$$\sum_{n \neq 0} \frac{(\psi_n(E))^2}{|n|^{1+\varepsilon}} < \infty, \tag{3.61}$$

and the proof is complete. □

We move now toward the characterization of the nature of the spectrum for H in the disordered case, namely (V_n) and IID sequence of random variable that are taking values in a bounded interval. V_n is of course defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, so also H is a random variable on this space: $\omega \mapsto H_\omega$. It is rather straightforward to verify the measurability of $\omega \mapsto H_\omega$, from Ω to the space of bounded linear operators equipped with the operator norm. The result that we are aiming at is that the spectrum of H is pure point, but we will not give a proof of the final result that is given in the next statement (a proof can be found in [**6**]). We will nevertheless make some steps toward the proof.

Theorem 3.13. *If the variance of V_1 is not zero, i.e. if V_1 is not constant, then $\Sigma(H) = \Sigma^p(H)$, $\Sigma^a(H) = \Sigma^s(H) = \emptyset$ and the eigenfunctions with eigenvalue E decay exponentially with rate $\gamma(E)$ (the Lyapunov exponent of the sequence of random matrices in* (3.5)*).*

So the spectrum of H is *made of* eigenvalues. Of course the union of all the eigenvalues are far from being the whole of $\Sigma(H)$, but they are dense in $\Sigma(H)$. In this sense the following result may at first look surprising:

Proposition 3.14. *Choose any $E \in \mathbb{R}$. Then the probability that E is an eigenvalue of H is zero.*

Proof. We consider the (self-adjoint) projection operator $\mathbf{1}_{\{E\}}(H)$. The trace of a projection is the dimension of the target space and if the target space if of dimension k then E is an eigenvalue of H of multiplicity k. So, by Remark 3.2,

$\mathrm{Trace}(\mathbf{1}_{\{E\}}(H)) \in \{0,1\}$. Therefore[1]

$$1 \ge \mathbb{E}\left[\mathrm{Trace}(\mathbf{1}_{\{E\}}(H))\right] = \mathbb{E}\sum_n \left(\mathbf{1}_{\{E\}}(H)e^n, e^n\right) = \sum_n \mathbb{E}\left[\mu_n(\{E\})\right]. \tag{3.62}$$

Translation invariance implies that $\mathbb{E}\left[\mu_n(\{E\})\right]$ does not depend on n and therefore $\mathbb{E}\left[\mu_n(\{E\})\right] = 0$. Therefore $\mu_{n,\omega}(\{E\}) = 0$ for every n, $\mathbb{P}(\mathrm{d}\omega)$-a.s.. Therefore $\mu_\omega(\{E\}) = \mu_{0,\omega}(\{E\}) + \mu_{-1,\omega}(\{E\}) = 0$ and (by Proposition 3.11), $\mathbb{P}(\mathrm{d}\omega)$-a.s., E is not an eigenvalue of H_ω. □

The next result approaches Theorem 3.13. It say that $\Sigma^{\mathrm{a}}(H) = \emptyset$. We state it in an equivalent way: recall the Lebesgue decomposition (3.44) that we now apply to $\mu = \mu_\omega$.

THEOREM 3.15. $\mu^{\mathrm{a}}_\omega(\mathbb{R}) = 0$, $\mathbb{P}(\mathrm{d}\omega)$-*a.s.*.

Actually, the proof says more, because it works with any fixed (i.e., non random) measure, it does not need to be the Lebesgue measure. More precisely, it says that given any measure ν on $\mathbb{R}$, $\mathbb{P}(\mathrm{d}\omega)$-a.s. there exists a Borel set B such that $\nu(\mathbb{R}\setminus B) = 0$ as well as $\mu^{\mathrm{a}}_\omega(B) = 0$. This is intriguing because we can choose $\nu = \mu_{\omega_0}$, highlighting thus the *wild dependence* of μ_ω on ω.

PROOF. Consider the product measure space of Ω equipped with the probability $\mathbb{P}$ and $\mathbb{R}$ with the Lebesgue measure. We set $S_n(\omega, E) = Y_n \ldots Y_1$, with Y_j defined in (3.5), and $\gamma(E)$ is the Lyapunov exponent associated to this sequence: we know that $\gamma(E) > 0$ for every E (see Proposition 3.4). Introduce the measurable subset of $\Omega \times \mathbb{R}$

$$W := \left\{(\omega, E) : \lim_{|n|\to\infty} \frac{1}{|n|} \log \|S_n(\omega, E)\| = \gamma(E) \text{ and } E \text{ is not eigenvalue of } H_\omega\right\}. \tag{3.63}$$

Note that for every $E \in \mathbb{R}$ the map (random variable) $\omega \mapsto \mathbf{1}_W(\omega, E)$ is almost surely equal to one, that is $\mathbb{P}(\{\omega : (\omega, E) \in W\}) = 1$. This is on one hand because of Theorem 1.8 (with (3.9) for the $n \to -\infty$ case) and, on the other, because of Proposition 3.14. Therefore for every $L > 0$ (in reality it sufficed to consider L such that $[-L, L]$ contains $\Sigma(H)$) by the Fubini-Tonelli Theorem

$$2L = \int_{[-L,L]} \mathbb{P}(\{\omega : (\omega, E) \in W\})\, \mathrm{d}E = \int_\Omega \left(\int_{[-L,L]} \mathbf{1}_W(\omega, E)\, \mathrm{d}E\right) \mathbb{P}(\mathrm{d}\omega). \tag{3.64}$$

But $\int_{[-L,L]} \mathbf{1}_W(\omega, E)\, \mathrm{d}E \le 2L$, hence $\mathbb{P}(\mathrm{d}\omega)$-a.s. $\int_{[-L,L]} \mathbf{1}_W(\omega, E)\, \mathrm{d}E = 2L$. We therefore consider for every $\omega \in \Omega$

$$A_{L,\omega} := \{E \in [-L, L] : (\omega, E) \in W\}, \tag{3.65}$$

which, $\mathbb{P}(\mathrm{d}\omega)$-a.s., differs from $[-L, L]$ only of a Lebesgue null set. But for $E \in A_{L,\omega}$ we know that if ψ solves $H_\omega \psi = E\psi$ then by Theorem A.3 (Oseledet's Theorem)

[1] In this step we use that $\mu_n(\{E\})$ is a random variable. This, i.e. measurability, is not obvious a priori: a proof can be found for example in [**1**, Lemma 2.1, pp. 206-207]

we have that ψ either tends to infinity or to zero, but in both cases exponentially fast. E is not an eigenvalue and therefore ψ tends to infinity exponentially fast for $n \to \infty$ or for $n \to -\infty$ (or both). Hence, by Proposition 3.12, $\mu_\omega(A_{L,\omega}) = 0$ (we recall that $\mu_\omega = \mu$ is the spectral measure $\mu_0 + \mu_{-1}$ of H_ω). Since $A_{L,\omega}$ is of full Lebesgue measure we see that μ_ω is singular with respect to the Lebesgue measure on $[-L, L]$ (in the sense, that it concentrates on a Lebesgue set of measure zero). The proof of Theorem 3.15 is therefore complete. $\square$

APPENDIX A

Birkhoff Ergodic Theorem

Let $(X, \mathcal{B}, \mu)$ a probability space and let $T : X \to X$ be a measure preserving transformation, i.e. T is measurable and $\mu(T^{-1}A) = \mu(A)$ for every $A \in \mathcal{B}$. One directly checks that $\mathcal{I} = \{A \in \mathcal{B} : \mu(T^{-1}A\Delta A) = 1\}$ is a σ-algebra. In particular, if T is ergodic, then $\mathcal{I}$ is trivial.

Theorem A.1 (Birkhoff Ergodic Theorem). *For every $f : X \to \mathbb{R}$ with $f \in \mathbb{L}^1(X, \mathcal{B}, \mu)$ there exists a real valued function $\bar{f}$, with $\bar{f} \in \mathbb{L}^1(X, \mathcal{B}, \mu)$, such that we have $\mu(\mathrm{d}x)$-a.s.*

$$\lim_{n\to\infty} \frac{f(x) + f(Tx) + \ldots + f(T^{n-1}x)}{n} = \bar{f}(x)\,, \tag{A.1}$$

and the convergence holds also in $\mathbb{L}^1(X, \mathcal{B}, \mu)$. Moreover $\bar{f}$ is the conditional expectation of f with respect to $\mathcal{I}$, that is $\bar{f}$ is measurable with respect to $\mathcal{I}$ and $\int_A f \,\mathrm{d}\mu = \int_A \bar{f} \,\mathrm{d}\mu$ for every $A \in \mathcal{I}$.

Proofs of the Birkhoff Ergodic Theorem may be found for example in [**2**, Ch. 6, Th. 6.1, pp. 113-115] and [**26**, Ch. 5, Sec. 3, pp. 409-411]. There only the a.s. convergence is stated. The $\mathbb{L}^1$ convergence can be extracted from the a.s. result by replacing f with $f_L = f\mathbf{1}_{[-L,L]} \in \mathbb{L}^\infty$, so if we call $S_{f,n}(x)$ the normalized sum in the left-hand side of (A.1), we have that $\lim_n S_{f_L,n} = \bar{f}_L$ a.s. and in $\mathbb{L}^1$ (by Dominated Convergence). Dominated Convergence also yields $\lim_{L\to\infty} \|f - f_L\|_1 = \lim_L \int_{\mathbb{R}} |f|\mathbf{1}_{|f|>L} \,\mathrm{d}\mu = 0$. Now it suffices to use

$$\left\|S_{f,n} - \bar{f}\right\|_1 \le \|S_{f,n} - S_{f_L,n}\|_1 + \left\|S_{f_L,n} - \bar{f}_L\right\|_1 + \left\|\bar{f}_L - \bar{f}\right\|_1\,, \tag{A.2}$$

and we have seen that the middle of the three terms in the right-hand side vanishes as $n \to \infty$ for every fixed L. By the measure preserving property and the triangular inequality we readily see that the first term is bounded by $\|f - f_L\|_1 \le \varepsilon$ for L sufficiently large. And Jensen inequality applied to the conditional expectation yields $\left\|\bar{f}_L - \bar{f}\right\|_1 \le \|f - f_L\|_1$ so also this term is bounded by ε for L large. Hence for every $\varepsilon > 0$ we have $\limsup_n \left\|S_{f,n} - \bar{f}\right\|_1 \le 2\varepsilon$ and we are done.

Birkhoff Ergodic Theorem holds also just assuming that $\int_{\mathbb{R}} f_+ \,\mathrm{d}\mu < \infty$: in this case of course the limit is not necessarily in $\mathbb{L}^1$ and one has to give up $\mathbb{L}^1$ convergence. Note that the conditional expectation of f with respect to $\mathcal{I}$ is well defined as the difference of the conditional expectation of f_+ (which is in $\mathbb{L}^1$) and of the conditional expectation of f_- (which is non-negative).

COROLLARY A.2. *For every $f : X \to \mathbb{R}$ with $\int_{\mathbb{R}} f_+ \, \mathrm{d}\mu < \infty$ we have that $\mu(\mathrm{d}x)$-a.s.*

$$\lim_{n\to\infty} \frac{f(x) + f(Tx) + \ldots + f(T^{n-1}x)}{n} \,=\, \bar{f}(x)\,, \tag{A.3}$$

where $\bar{f}$ is the conditional expectation of f with respect to $\mathcal{I}$.

PROOF. With $E_\mu[\cdot]$ for the expectation on $(X, \mathcal{B}, \mu)$, in particular $\bar{f} = E_\mu[f|\mathcal{I}]$, we introduce for $L > 0$ the event $A_L := \{x : E_\mu[f|\mathcal{I}](x) > -L\} \in \mathcal{I}$. So for $x \in A_L$ we have $E_\mu[f_-|\mathcal{I}](x) < L + E_\mu[f_+|\mathcal{I}](x)$, hence $E_\mu[f_- \mathbf{1}_{A_L}] = E_\mu[E_\mu[f_-|\mathcal{I}]\mathbf{1}_{A_L}] < L + E_\mu[f_+]$. Therefore $g := f\mathbf{1}_{A_L} \in \mathbb{L}^1$ and by Theorem A.1 we have that a.s. $\lim_n S_{g,n} = E[g|\mathcal{I}] = E[f|\mathcal{I}]\mathbf{1}_{A_L}$. On the other hand, since $A_L \in \mathcal{I}$, we readily see that $\mu(\mathrm{d}x)$-a.s.

$$S_{g,n}(x) \,=\, \frac{f(x) + f(Tx) + \ldots + f(T^{n-1}x)}{n}\mathbf{1}_{A_L}(x)\,, \tag{A.4}$$

and therefore, with $A_\infty := \cup_L A_L$, we have that $\mu(\mathrm{d}x)$-a.s.

$$\lim_{n\to\infty} \frac{f(x) + f(Tx) + \ldots + f(T^{n-1}x)}{n}\mathbf{1}_{A_\infty}(x) \,=\, E_\mu[f|\mathcal{I}](x)\mathbf{1}_{A_\infty}(x)\,. \tag{A.5}$$

On the other hand, for $x \in A_\infty^\complement$ we have $E_\mu[f|\mathcal{I}](x) = -\infty$ (equivalently, $E_\mu[f_-|\mathcal{I}](x) = \infty$), so we are left with showing that a.s $\lim_n S_{f,n}\mathbf{1}_{A_\infty^\complement} = -\infty\mathbf{1}_{A_\infty^\complement}$ or, equivalently, that $\lim_n S_{f_-,n}\mathbf{1}_{A_\infty^\complement} = \infty\mathbf{1}_{A_\infty^\complement}$. By using that $A_\infty^\complement \in \mathcal{I}$ we see that it suffices to show that a.s. $\lim_n S_{g,n} = \infty\mathbf{1}_{A_\infty^\complement}$ where this time we used $g := f_-\mathbf{1}_{A_\infty^\complement}$. But what Theorem A.1 tells us is that for every $L > 0$ a.s.

$$\liminf_n S_{g,n} \,=\, \lim_n S_{g\wedge L,n} \,=\, E_\mu\left[g \wedge L \,|\, \mathcal{I}\right] \,=\, E_\mu\left[f_- \wedge L \,|\, \mathcal{I}\right]\mathbf{1}_{A_\infty^\complement}\,. \tag{A.6}$$

By the Monotone Convergence we pass to the $L \to \infty$ limit and the rightmost term becomes $\infty\mathbf{1}_{A_\infty^\complement}$, so the proof of the corollary is complete. □

The following form of the result known as *Oseledets Theorem* is stated in [**1**, Prop. 1.1, p. 188-189]. It does not involve any probability, just linear algebra. The proof given here is a mild adaptation of the one in [**29**].

THEOREM A.3. *Let $Y_1, Y_2, \ldots$ be two by two real matrices with absolute value of determinant equal to 1 and such that*

(1) $\lim_n (1/n) \log \|Y_n Y_{n-1} \ldots Y_1\| = \gamma > 0$;

(2) $\lim_n (\log \|Y_n\|)/n = 0$.

Then there exists a vector $V \neq 0$ such that

$$\lim_n \frac{1}{n} \log \|Y_n Y_{n-1} \ldots Y_1 V\| \,=\, -\gamma\,, \tag{A.7}$$

and for every vector $U \neq 0$ which is not collinear with V, i.e. $\overline{U} \neq \overline{V}$, we have

$$\lim_n \frac{1}{n} \log \|Y_n Y_{n-1} \dots Y_1 U\| \,=\, \gamma\,. \tag{A.8}$$

Proof. We set $M_n := Y_n Y_{n-1} \dots Y_1$. Since $\|M_n\|$, which we may assume to be larger than 1 without loss of generality, is the square root of the larger of the two eigenvalues of $M_n^* M_n$ and since the product of the two eigenvalues is one, we see that, if we call U_n the eigenvector corresponding to the larger eigenvalue and V_n the other eigenvector (we choose $\|V_n\| = \|U_n\| = 1$), we have

$$\|M_n U_n\| \,=\, \|M_n\| \quad \text{ and } \quad \|M_n V_n\| \,=\, 1/\|M_n\|\,. \tag{A.9}$$

Let us stress that V_n and U_n are orthogonal. Moreover note that the normalization does not identify U_n and V_n, but rather $\pm U_n$ and $\pm V_n$. It is about this *sign* that we are going to refer to in the rest of the proof when saying *by a suitable choice of the signs.*

Recall that, for $x, y \in \mathbb{R} \setminus \{0\}$ we use on the projective space the distance $\mathbf{d}(\overline{x}, \overline{y})$ which is the absolute value of the sine of the angle between the two rays. For ease of notation we write $\mathbf{d}(x, y)$ for $\mathbf{d}(\overline{x}, \overline{y})$.

The crucial claim is that

$$\limsup_n \frac{1}{n} \log \mathbf{d}\left(V_n, V_{n+1}\right) \,\leq\, -2\gamma\,. \tag{A.10}$$

In fact, from (A.10) one easily obtains that the convergence $\lim_n V_n =: V$ in the projective space, or, equivalently, $\lim_n V_n = V$ in $\mathbb{R}^2$ with a suitable choice of the signs. This is because (V_n) is a Cauchy sequence: $\mathbf{d}\left(V_n, V_{n+1}\right) \leq \exp(-cn)$ for any $c < 2\gamma$ and n sufficiently large. So there exists $C > 0$ such that $\mathbf{d}\left(V_n, V_{n+m}\right) \leq C \exp(-cn)$ for every $m \in \mathbb{N}$. This also implies that $\mathbf{d}\left(V_n, V\right) \leq C \exp(-cn)$, in particular

$$\limsup_n \frac{1}{n} \log \mathbf{d}\left(V_n, V\right) \,\leq\, -2\gamma\,. \tag{A.11}$$

Let us give a proof of (A.10). For this we call θ_n the angle angle between V_n and V_{n+1}. We can write V_n as linear combination of U_{n+1} and V_{n+1}:

$$V_n \,=\, \cos(\theta_n) V_{n+1} + \sin(\theta_n) U_{n+1}\,. \tag{A.12}$$

By using the fact that also $M_{n+1} V_{n+1} \perp M_{n+1} U_{n+1}$ we obtain

$$\|M_{n+1} V_n\| \,\geq\, |\sin(\theta_n)| \|M_{n+1} U_{n+1}\| \,\geq\, |\sin(\theta_n)| \|M_{n+1}\|\,, \tag{A.13}$$

and

$$\|M_{n+1} V_n\| \,\leq\, \|Y_{n+1}\| \|M_n V_n\| \,=\, \|Y_{n+1}\| / \|M_n\|\,. \tag{A.14}$$

Therefore

$$|\sin(\theta_n)| \,\leq\, \frac{\|Y_{n+1}\|}{\|M_n\| \|M_{n+1}\|}\,, \tag{A.15}$$

and by exploiting both the assumptions (1) and (2) of Theorem A.3 we see that the superior limit of $(1/n) \log |\sin(\theta_n)|$ is not larger than -2γ and (A.10) is proven.

The next step is showing that V is contracted by M_n at the same exponential rate as V_n, that is

$$\lim_n \frac{1}{n} \log \|M_n V\| \;=\; -\gamma\,, \tag{A.16}$$

which is (A.7). To show this we now redefine θ_n to be the angle between V and V_n, so $V = \cos(\theta_n)V_n + \sin(\theta_n)U_n$ (here again, a choice of the signs is made) and by exploiting once again the orthogonality of M_nV_n and M_nU_n and $\sqrt{x_1^2 + x_2^2} \le |x_1| + |x_2|$ we see that

$$\begin{aligned} \|M_n V\| \;&\le\; |\cos(\theta_n)|\|M_n V_n\| + |\sin(\theta_n)|\|M_n U_n\| \\ &=\; \frac{|\cos(\theta_n)|}{\|M_n\|} + |\sin(\theta_n)|\|M_n\| \\ &\le\; \frac{1}{\|M_n\|} + \mathbf{d}(V, V_n)\|M_n\| \;\le\; e^{-(\gamma-\varepsilon)n} + e^{-2(\gamma-\varepsilon)n} e^{(\gamma+\varepsilon)n}\,, \end{aligned} \tag{A.17}$$

where we have used $|\sin(\theta_n)| = \mathbf{d}(V, V_n)$ and the last step, which holds for every $\varepsilon > 0$ and for n sufficiently large, follows from (A.11) and the hypothesis on the asymptotic behavior of $\|M_n\|$. Therefore (A.16) is established.

We are therefore left with showing (A.8). But for this is sufficient to remark, since $V_n \perp U_n$, also (U_n) converges (in the projective space) to $V^\perp$, at the same exponential speed (even if we will not use that). Therefore if we write $V^\perp = aV_n + bU_n$, for n large we have $|b|$ close to one and a close to zero. Therefore for n large $|a|$ is bounded away from zero and this suffices to conclude that $\lim_n (1/n) \log \|M_n V^\perp\| = \gamma$. Since the vector U that appears in (A.8) can also be written as $aV + bV^\perp$ with $b \neq 0$, the proof is complete. □

APPENDIX B

Regularity of Lyapunov exponents

Let us start with an example (due to H. Furstenberg and Y. Kiefer [**16**] that shows that the Lyapunov exponent is, in general, not a continuous function of the law of the random matrices. Consider in fact the Lyapunov exponent $\gamma(p)$ of the product of the random matrices $(M_1 X_j + M_2(1-X_j))$, with (X_j) IID Bernoulli random variables of parameter p and

$$M_1 := \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \quad \text{and} \quad M_2 := \begin{pmatrix} 0 & b \\ 1/b & 0 \end{pmatrix}, \tag{B.1}$$

and $a > 1$ and $b > 0$ two constants. This is not a nice random matrix sequence because it does not satisfy irrreducibility: in fact the union of the x and y axes is stable under the action of the group generated by M_1 and M_2. On the other hand, it is not difficult to compute $\gamma(p)$ and see that

Proposition B.1. *$\gamma(p) = 0$ for $p \in [0,1)$.*

On the other hand, it is evident that $\gamma(1) = \log a > 0$. Hence $p \mapsto \gamma(p)$ has a discontinuity at 1.

Proof. For the case $p = 0$ the result is straightforward because M_2^2 is the identity matrix. Let us the assume that $p \in (0,1)$: the Markov chain underlying this matrix product is very simple (see Fig. B.1) if we start from the unit vectors of the two axes (and of course we can recover the general case by superposition, so let us focus on them). If one starts from $(1,0)^*$ with probability p this direction is preserved and the length is multiplied by a. With probability $1-p$ instead we step to $(0,1)^*$, and the vector is multiplied by b. The dynamics is similar from this other state, except that a is replaced by $1/a$ and b by $1/b$. Therefore if introduce the sequence of random times $\tau_1 = \inf\{n : X_j = 0\}$ then one easily infers that

$$M_{\tau_{2j}} M_{\tau_{2j}-1} \ldots M_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a^{\sum_{k=1}^{j} Y_k} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{B.2}$$

where (Y_j) is a sequence of IID random variables and $Y_1 \sim G - G'$, where G and G' are two independent Geometric random variables of parameter p, i.e. the probability that G is $n = 0, 1, \ldots$ is $p^n(1-p)$. In the same way we see that

$$M_{\tau_{2j}} M_{\tau_{2j}-1} \ldots M_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = a^{-\sum_{k=1}^{j} Y_k} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{B.3}$$

But $\sum_{k=1}^{j} Y_k/\sqrt{k}$ converges in law toward a Gaussian random variable and one can apply an iterated logarithm result to get to a sharp estimate. Here we just content ourselves with the Law of Large Numbers that tells us that $\lim_j \sum_{k=1}^{j} Y_k/k = 0$ a.s. and in $\mathbb{L}^1$ that is telling us that

$$\frac{1}{n}\log\left((1,0)M_{\tau_{2j}}M_{\tau_{2j}-1}\ldots M_1\begin{pmatrix}1\\0\end{pmatrix}\right) = 0\,, \tag{B.4}$$

a.s. and in $\mathbb{L}^1$. The very same result holds with $(1,0)$ replaced by $(0,1)$. Since the case of cross products clearly yield zero, we readily see that we can also replace $(1,0)$ with $(1,1)$ and this is a norm of the matrix product because we are multiplying matrices with non-negative entries.

Therefore the proof of Proposition B.1 is complete. □

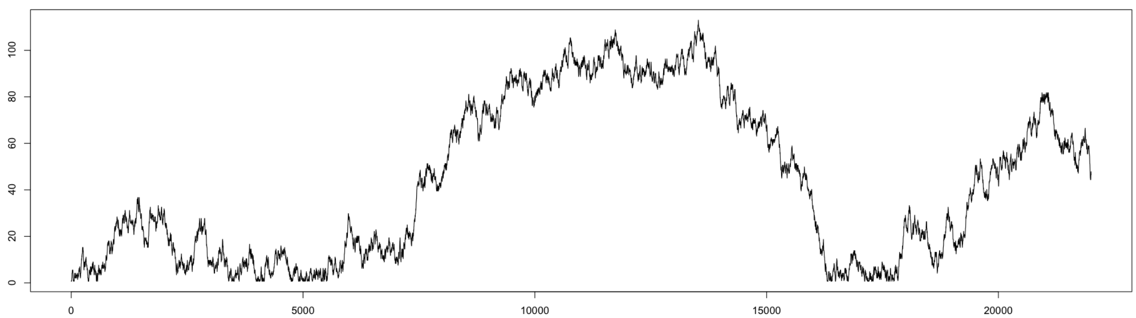

FIGURE B.1. The plot of $n \mapsto \log \|Y_n \cdot Y_1\|_1 = \log \sum_{i,j=1}^{2} (Y_n \cdot Y_1)_{i,j}$, with (Y_n) as in Proposition B.1, $p = 1/2$, $a = 2$ and $b = 1$. As expected from the argument of proof, the trajectory is close to being a random walk (or Brownian motion) reflected at or near 0. More precisely the minimum of $\|Y_n \cdot Y_1\|_1$ is $\log(2)$ so the reflection is at this level. Therefore the Lyapunov exponent is zero. Note also that the behavior of the four entries entries of $Y_n \cdot Y_1$ is very different: $Y_n \cdot Y_1$ is either a diagonal matrix (this happens in particular if $Y_n = M_1$) or an anti-diagonal matrix (this happens in particular if $Y_n = M_2$ and $Y_{n-1} = M_1$): in this case $\lim_n (1/n)(Y_n \cdot Y_1)_{i,j})$ does not exist with probability one for every choice of $(i,j) \in \{1,2\}^2$, simply because $(Y_n \cdot Y_1)_{i,j} = 0$ for infinitely many values of n.

Here is a general result that shows that some regularity (upper semi-continuity) does hold under rather minimal conditions. We write $\gamma(\mu)$ for the Lyapunov exponent of (Y_n) IID and μ is the law of Y_1.

PROPOSITION B.2. *For every sequence (μ_n) of probabilities on $\mathbb{G}$ such that, using $Y^{(n)}$ for a random variable with law μ_n, we have*

(1) $(\log_+ \|Y^{(n)}\|)$ is uniformly integrable;

(2) $(Y^{(n)})$ converges in law toward $Y \sim \mu$;

then

$$\limsup_{n\to\infty} \gamma(\mu_n) \le \gamma(\mu)\,. \tag{B.5}$$

Proof. With $S_N^{(n)} = Y_N^{(n)} \cdot Y_1^{(n)}$ and $S_N = Y_N \cdot Y_1$ we have by sub-additivity that for every $N \in \mathbb{N}$ and every $L \in \mathbb{R}$

$$\gamma(\mu_n) \le \frac{1}{N}\mathbb{E}\left[\log \|S_N^{(n)}\|\right] \le \mathbb{E}\left[\left(\frac{1}{N}\log\|S_N^{(n)}\|\right) \vee L\right]. \tag{B.6}$$

Assume now that $L \le 0$. Then $L \le ((1/N)\log\|S_N^{(n)}\|) \vee L \le (1/N)\sum_{j=1}^N \log_+ \|Y_j^{(n)}\|$. Therefore, given N and L, the sequence of random variables $(((1/N)\log\|S_N^{(n)}\|) \vee L)_{n\in\mathbb{N}}$ is uniformly integrable and converges in law to $((1/N)\log\|S_N\|) \vee L$. Therefore

$$\limsup_{n\to\infty} \gamma(\mu_n) \le \mathbb{E}\left[\left(\frac{1}{N}\log\|S_N\|\right) \vee L\right], \tag{B.7}$$

and, by using that $\lim_N (1/N)\log\|S_N\| = \gamma(\mu)$ a.s. (Theorem 1.5) and that $\log_+ S_N \le \sum_{j=1}^N \log_+ \|Y_j\|$ from which we recover uniform integrability, we obtain

$$\limsup_{n\to\infty} \gamma(\mu_n) \le \gamma(\mu) \vee L, \tag{B.8}$$

for every negative L. The proof is therefore complete. □

Continuity holds if we are under the hypotheses of Theorem 1.8, plus some uniform integrability:

Proposition B.3. *With the same notations of Proposition B.2, if*

1. *the probability that* $\det(Y^{(n)}) = 1$ *for every* n;
2. $Y^{(n)}$ *converges in law toward* $Y \sim \mu$ *and* Y *satisfies the assumptions of Theorem 1.8 (i.e., non compactness and irreducibility);*
3. $(\log_+ \|Y^{(n)}\|)$ *is uniformly integrable;*

then

$$\lim_{n\to\infty} \gamma(\mu_n) = \gamma(\mu). \tag{B.9}$$

Proof. Let us remark that if μ satisfies the properties (a) and (b) in Theorem 1.8, i.e. non compactness and irreducibility, these two properties are satisfied also in a neighborhood of μ: without loss of generality we assume that they are satisfied by μ_n for every n. By recalling that the integrability is given by hypothesis (3), we can therefore use the Furstenberg formula (Theorem 1.8)

$$\gamma(\mu_n) = \int_{\mathbb{G}}\int_{\mathbb{B}} \log\left(\frac{\|Mx\|}{\|x\|}\right)\mu_n(\mathrm{d}M)\nu_n(\mathrm{d}\overline{x}), \tag{B.10}$$

with ν_n the unique probability on $\mathbb{B}$ that solves $\mu_n \star \nu_n = \nu_n$. The analogous formula holds also for the limit case, i.e. for $\gamma(\mu)$. Then, byy compactness of $\mathbb{B}$, we can extract convergent subsequences (ν_{n_j}): we remark that any limit probability ν is μ invariant because $\mu_n \star \nu_n = \nu_n$ is equivalent to $\int_{\mathbb{G}}\int_{\mathbb{B}} f(M\cdot\overline{x})\mu_n(\mathrm{d}M)\nu_n(\mathrm{d}\overline{x}) = \int_{\mathbb{B}} f(\overline{x})\nu_n(\mathrm{d}\overline{x})$ for every $f \in C^0(\mathbb{B};\mathbb{R}) = C_b^0(\mathbb{B};\mathbb{R})$ and $(\mu_{n_j} \otimes \nu_{n_j})$ converges to $\mu \otimes \nu$, so $\int_{\mathbb{G}}\int_{\mathbb{B}} f(M\cdot\overline{x})\mu(\mathrm{d}M)\nu(\mathrm{d}\overline{x}) = \int_{\mathbb{B}} f(\overline{x})\nu(\mathrm{d}\overline{x})$. Since there exists a unique probability ν satisfying $\mu \star \nu = \nu$ (Theorem 1.8), the sequence (ν_n) converges to ν.

In order to show that $\lim_n \gamma(\mu_n) = \gamma(\mu)$ it suffices therefore to remark the elementary fact that $(M, \overline{x}) \mapsto \sigma(M, \overline{x}) := \log(\|Mx\|/\|x\|)$ is C^0 and (less straightforward) the uniform integrability

$$\lim_{L\to\infty} \int_{\mathbb{G}} \int_{\mathbb{B}} |\sigma(M, \overline{x})| \, \mathbf{1}_{\{|\sigma(M,\overline{x})|>L\}}(M, \overline{x}) \mu_n(\,\mathrm{d}M) \nu_n(\,\mathrm{d}\overline{x}) \;=\; 0\,. \tag{B.11}$$

For this we recall that $\sigma(M, \overline{x}) \le \log \|M\|$ for every $\overline{x}$ and that

$$\sigma(M, \overline{x}) \;=\; -\log\left(\frac{\|M^{-1}Mx\|}{\|Mx\|}\right) \;\ge\; -\log\|M^{-1}\|\,. \tag{B.12}$$

But $\det(M) = 1$ implies that $\|M\| = \|M^{-1}\|(\ge 1)$: in fact

$$\|M^{-1}\|^2 = \lambda_{\max}\left(\left(M^{-1}\right) * M^{-1}\right) = \lambda_{\max}\left((MM^*)^{-1}\right) = \\ \frac{1}{\lambda_{\min}\left(MM^*\right)} = \lambda_{\max}\left(MM^*\right) = \lambda_{\max}\left(M^*M\right) = \|M\|^2\,. \tag{B.13}$$

We conclude that $|\sigma(M, \overline{x})| \le \log\|M\| = \log_+\|M\|$ so (B.11) follows from hypothesis (3). The proof of Proposition B.3 is therefore complete. □

REMARK B.4. *Here is an interesting example: with $a > 1$ and $\theta \in \mathbb{R}$*

$$A_1 \;:=\; \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \quad \textit{and} \quad R_\theta \;:=\; \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}\,, \tag{B.14}$$

we set $A_2 = R_\theta A_1 R_{-\theta}$ and $Y_j := X_j A_1 + (1 - X_j) A_2$, (X_j) IID $B(1/2)$. Then by Proposition B.3 we see that the top Lyapunov exponent $\gamma(a, \theta)$ of (Y_j) is continuous in $(a, \theta) \in (1, \infty) \times (\mathbb{R} \setminus (\pi/2)\mathbb{Z})$: note in particular that it is easy to see that $\gamma(a, \pi/2 + n\pi) = 0$ for every n and that it is straightforward to see that $\gamma(a, n\pi) = \log a$. In fact (see [**29**, Th. 1.3 and Example 1.4]*), $\gamma(a, \theta)$ is continuous at every $(a, \theta) \in (1, \infty) \times \mathbb{R}$. See Fig. B.2 for a numerical plot of $\theta \mapsto \gamma(2, \theta)$.*

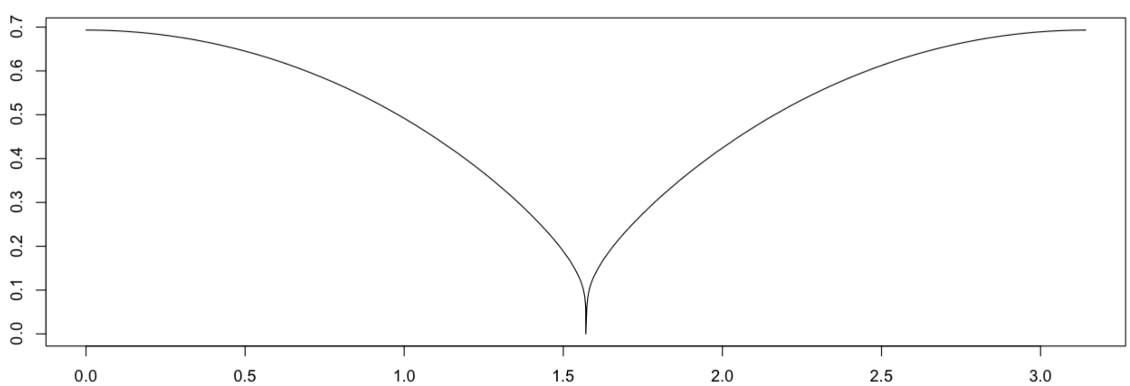

FIGURE B.2. numerical plot of $\theta \mapsto \gamma(2, \theta)$ of Remark B.4.

Next we state a somewhat general result, which is a very special case of the general result due to D. Ruelle [**25**]. We work as usual in $\mathbb{R}^2$: we need the notion of a proper convex cone (we will simply say: cone) which in $\mathbb{R}^2$ is simply a subset $\{0\} \cup \{y \in \mathbb{R} \setminus \{0\} : \mathbf{d}(\overline{x}, \overline{y}) \le 1 - \delta\}$ for an $x \neq 0$ and a $\delta \in (0,1)$.

Theorem B.5. *We consider a sequence $((Y_n, M_n))$ of IID $GL_2(\mathbb{R}) \times \mathcal{M}_2(\mathbb{R})$ random variables. We assume that*

(1) $\log \|Y_1\| \in \mathbb{L}^1$;
(2) there exists two cones C_1 and C_2, with C_2 strictly contained in C_1 such that if $x \in \mathrm{C}_1$, then $\mathbb{P}(Y_1 x \in \mathrm{C}_2) = 1$;
(3) there exists $c > 0$ such that $\mathbb{P}(\|M_1\|/\|Y_1\| \le c) = 1$;

then the top Lyapunov exponent $\gamma(\delta)$ associated to $(Y_n + \delta M_n)$ is well defined for every $\delta \in \mathbb{R}$. Moreover $\delta \mapsto \gamma(\delta)$ is real analytic for δ in a neighborhood of 0.

Proof. The conditions yield that $\|Y_1 + \delta M_1\| \le (1 + |\delta| c)\|Y_1\|$, hence $\mathbb{E}[\log_+ \|Y_n + \delta M_n\|] < \infty$. This is sufficient to define $\gamma(\delta)$ (Definition 1.3). For the rest we apply [**8**, Theorem 1.3]: note in particular that for δ small we have that $\mathbb{E}[|\log \|Y_n + \delta M_n\||] < \infty$. $\square$

As an application we consider the Lyapunov exponent $\gamma(\varepsilon)$ associated to (T_j) in (2.27):

$$T_j = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon Z_j & Z_j \end{pmatrix}, \tag{B.15}$$

with (Z_j) IID and $\log Z_1 \in \mathbb{L}^1$. We have that:

Corollary B.6. *With the definitions we have just introduced, $\varepsilon \mapsto \gamma(\varepsilon)$ is real analytic in $\varepsilon \in (-1,1) \setminus \{0\}$.*

Proof. For this it suffices to choose an arbitrary $\varepsilon \neq 0$ (without loss of generality we choose $\varepsilon > 0$). We are going to work with $Y_j = T_j^*$ in Theorem B.5 and

$$M_j = \begin{pmatrix} 0 & Z_j \\ 1 & 0 \end{pmatrix}. \tag{B.16}$$

We are exploiting here the fact that the Lyapunov exponent is the same if we consider the transpose of the matrices.

Hypothesis (1) is obvious. Let us focus on (2) and (3).

For (2) we choose $\mathrm{C}_1 = \{x : x_1 \ge 0 \text{ and } x_2 \ge 0\} \cup \{x : x_1 \le 0 \text{ and } x_2 \le 0\}$, i.e. the first and third quadrant. The we observed that the action of $Y = Y_j$ sends $r := x_2/x_1$ to $(\varepsilon + Zr)/(1 + \varepsilon Zr)$ and for every $z, r \in (0, \infty)$

$$\varepsilon \le \varepsilon \left(1 + \frac{zr(1-\varepsilon^2)}{\varepsilon(1+\varepsilon zr)}\right) = \frac{\varepsilon + zr}{1 + \varepsilon zr} = \frac{1}{\varepsilon}\left(1 - \frac{1-\varepsilon^2}{1+\varepsilon zr}\right) \le \frac{1}{\varepsilon}, \tag{B.17}$$

so the cone C_1 is contracted into the cone that is between the line with angle $\arctan(\varepsilon)$ with the x axis and the line with angle with angle $\arctan(1/\varepsilon)$ with the same axis.

For (3) a straightforward computation yields

$$\begin{aligned} \|T\|^2 &= \frac{1}{2}\left((1+\varepsilon^2)(1+Z^2) + \sqrt{(1+\varepsilon^2)^2(1+Z^2)^2 - 4Z^2(1-\varepsilon^2)^2}\right) \\ &\geq \frac{1}{2}(1+\varepsilon^2)(1+Z^2)\,, \end{aligned} \tag{B.18}$$

and $\|M\| = Z$. So $\|M\|/\|T\| = \|M\|/\|T^*\| \leq 2$. □

APPENDIX C

Some statistical mechanics

We start by working on a finite set Λ and $H : \{-1,1\}^\Lambda \to \mathbb{R}$. We set for $\beta \geq 0$

$$\mu_\beta(\sigma) := \frac{\exp(\beta H(\sigma))}{Z_{\Lambda,\beta}} \quad \text{with} \quad Z_{\Lambda,\beta} := \sum_\sigma \exp(\beta H(\sigma)) \,. \tag{C.1}$$

The free energy density is

$$\mathrm{F}_\Lambda(\beta) := \frac{1}{|\Lambda|} \log Z_{\Lambda,\beta} \,. \tag{C.2}$$

By Jensen inequality for every probability ν on $\{-1,1\}^\Lambda$

$$\log Z_{\Lambda,\beta} = \log \sum_\sigma \nu(\sigma) \frac{\exp(\beta H(\sigma))}{\nu(\sigma)} \leq \beta \sum_\sigma \nu(\sigma) H(\sigma) - \sum_\sigma \nu(\sigma) \log \nu(\sigma) \,, \tag{C.3}$$

from which we obtain the variational formula

$$\mathrm{F}_\Lambda(\beta) = \min_\nu \left(\beta \sum_\sigma \nu(\sigma) \frac{H(\sigma)}{|\Lambda|} - \frac{1}{|\Lambda|} \sum_\sigma \nu(\sigma) \log \nu(\sigma) \right) , \tag{C.4}$$

because equality holds for $\nu = \mu_\beta$:

$$\mathrm{F}_\Lambda(\beta) = \left(\beta \sum_\sigma \mu_\beta(\sigma) \frac{H(\sigma)}{|\Lambda|} - \frac{1}{|\Lambda|} \sum_\sigma \mu_\beta(\sigma) \log \mu_\beta(\sigma) \right) . \tag{C.5}$$

Let us introduce a formalism to make the partition function more explicit: for $n < m$ and $\sigma_\ell, \sigma_\imath \in \{-1,1\}$ we introduce

$$Z_{n,m,\sigma_\ell,\sigma_\imath} := \sum_{\sigma \in \{-1,1\}^{\{n,\ldots,m+1\}}} \exp\left(J \sum_{j=n+1}^{m} \sigma_{j-1}\sigma_j + \sum_{j=n}^{m} h_j \sigma_j \right) \tag{C.6}$$

where $\sigma_n = \sigma_\ell$ and $\sigma_{m+1} = \sigma_\imath$ are the boundary conditions. We identify in what follows the spin up $\sigma = +1$ with the column vector $(1,0)^*$ and the spin down $\sigma = -1$ with the column vector $(0,1)^*$. So

$$Z_{n,m,\sigma_\ell,\sigma_\imath} = \sigma_\ell^* T_{n+1} \ldots T_m \sigma_\imath \,, \tag{C.7}$$

with

$$T_j = \begin{pmatrix} e^{J+h_j} & e^{-J+h_j} \\ e^{-J-h_j} & e^{J-h_j} \end{pmatrix} . \tag{C.8}$$

Let us apply (C.5) to the one dimensional Ising case with $h = 0$. In this case the matrix is

$$\begin{pmatrix} e^J & e^{-J} \\ e^{-J} & e^J \end{pmatrix} = e^J \left(1 + e^{-2J}\right) \begin{pmatrix} \frac{1}{1+e^{-2J}} & \frac{e^{-2J}}{1+e^{-2J}} \\ \frac{e^{-2J}}{1+e^{-2J}} & \frac{1}{1+e^{-2J}} \end{pmatrix} . \tag{C.9}$$

With $p = e^{-2J}/(1 + e^{-2J})$ we see that the entropy contribution gives for renewal with inter-arrival $K(\cdot)$ and then for $K(n) = (1-p)^{n-1}p$

$$\frac{\sum_n K(n) \log K(n)}{\sum_n nK(n)} = p \log p + (1-p) \log(1-p) = p \log p - p + O(p^2) . \tag{C.10}$$

This means that the free energy is for $J \to \infty$

$$\begin{aligned} J + 2J \log\left(\frac{e^{-2J}}{1+e^{-2J}}\right) - \frac{e^{-2J}}{1+e^{-2J}} \log\left(\frac{e^{-2J}}{1+e^{-2J}}\right) + \frac{e^{-2J}}{1+e^{-2J}} + O\left(e^{-4J}\right) \\ = J + e^{-2J} + O\left(e^{-4J}\right) . \end{aligned} \tag{C.11}$$

Now we repeat the computation for a more generic $K(\cdot)$ for which $\sum_n nK(n) \sim CJ^2$. For the free energy we obtain

$$J - \frac{2J}{CJ^2(1 + o(1))} - \frac{\sum_n K(n) \log K(n)}{CJ^2(1 + o(1))} , \tag{C.12}$$

so if $K(n) = (1-p)^{n-1}p$, then $p \sim 1/(CJ^2)$

Bibliography

[1] P. Bougerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, Progress in Probability and Statistics, **8**. Birkhäuser, 1985.

[2] L. Breiman, *Probability*, Corrected reprint of the 1968 original, Classics in Applied Mathematics, **7**. SIAM, 1992.

[3] D. Buraczewski, E. Damek and T. Mikosch, *Stochastic models with power-law tails. The equation X=AX+B*, Springer Series in Operations Research and Financial Engineering. Springer, 2016.

[4] A. Crisanti, G. Paladin and A. Vulpiani, *Products of random matrices in statistical physics*, volume 104 of *Springer Series in Solid-State Sciences*. Berlin: Springer-Verlag, 1993.

[5] C. de Calan, J. M. Luck, Th. M. Nieuwenhuizen and D. Petritis, *On the distribution of a random variable occurring in 1D disordered systems*, J. Phys. A **18** (1985), 501-523.

[6] R. Carmona, A. Klein and F. Martinelli, *Anderson localization for Bernoulli and other singular potentials*, Comm. Math. Phys. **108** (1987), 41-66.

[7] F. Comets, G. Giacomin and R. L. Greenblatt, *Continuum limit of random matrix products in statistical mechanics of disordered systems*, Commun. Math. Phys. **369** (2019), 171-219.

[8] L. Dubois, *Real cone contractions and analyticity properties of the characteristic exponents*, Nonlinearity **21** (2008), 2519-2536.

[9] B. Derrida and H. J. Hilhorst, *Singular behaviour of certain infinite products of random* 2×2 *matrices*. J. Phys. A **16** (1983), 2641-2654.

[10] S. Ethier and T. Kurtz, *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. 1986.

[11] H. Furstenberg, *Noncommuting random products* Trans. Amer. Math. Soc. **108** (1963), 377-428.

[12] H. Furstenberg and H. Kesten, *Products of random matrices*, Ann. Math. Statist. **31** (1960), 457-469.

[13] B. Havret, *Regular expansion for the characteristic exponent of a product of* 2×2 *random matrices*, Math. Phys. Anal. Geom. **22** (2019), no. 2, Paper No. 15, 30 pp..

[14] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems.* Third edition. Grundlehren der Mathematischen Wissenschaften, **260**, Springer, 2012.

[15] H. L. Frisch and S. P. Lloyd, *Electron levels in a one-dimensional random lattice*, Phys. Rev. **120** (1960), 1175-1189.

[16] H. Furstenberg and Y. Kifer, *Random matrix products and measures on projective spaces*, Israel J. Math. **46** (1983), 12-32.

[17] G. Genovese, G. Giacomin and R. L. Greenblatt, *Singular behavior of the leading Lyapunov exponent of a product of random 2×2 matrices*, Commun. Math. Phys. **351** (2017), 923-958.

[18] G. Giacomin and R. L. Greenblatt, *Lyapunov exponent for products of random Ising transfer matrices: the balanced disorder case*, ALEA-Latin American Journal of Probability **19** (2022), 701-728.

[19] T. Kato, *Perturbation Theory of Linear Operators*, Classics in Mathematics **132**, reprint of the 1980 edition, Springer, 1995.

[20] H. Kesten, *Random difference equations and renewal theory for products of random matrices*, Acta Math. **131** (1973), 207-248.

[21] G. Maruyama and H. Tanaka, *Ergodic property of N-dimensional recurrent Markov processes.* Mem. Fac. Sci. Kyushu Univ. Ser. A **13** (1959), 157-172.

[22] B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model.* Harvard University Press, 1973.

[23] B. M. McCoy and T. T. Wu, *Theory of a two-dimensional Ising model with random impurities. I. Thermodynamics.* Phys. Rev. **176** (1968), 631-643.

[24] F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds., *DLMF. NIST Digital Library of Mathematical Functions*, http://dlmf.nist.gov/, Release 1.0.28 of 2020-09-15 (2020).

[25] D. Ruelle, *Analycity properties of the characteristic exponents of random matrix products. Adv. Math* **32** (1979), 68-80.

[26] A. N. Shiryaev, *Probability*, Second edition, Graduate Texts in Mathematics, **95**. Springer-Verlag, 1996.

[27] D. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes.* Classics in Mathematics. Springer-Verlag, Berlin, 2006.

[28] G. Teschl, *Mathematical methods in quantum mechanics. With applications to Schrödinger operators.* AMS, 2009. Available from http://www.mat.univie.ac.at/ gerald/

[29] M. Viana, *Lectures on Lyapunov exponents*, Cambridge Studies in Advanced Mathematics, **145** Cambridge University Press, 2014.