Stochastic Dynamics

Lecture Course
Peter Imkeller

Humboldt Universität zu Berlin

WS 2015/16

manuscript by Wolfgang Siegert

Contents

1. Markov chains: construction and elementary properties

Definition 1.1. Let $(S,\mathscr{S})$ be a measurable space. A function

$$p : S \times \mathscr{S} \to [0,1]$$

ia called *transition probability*, if

(a) for each $x \in S$ $p(x,\,.\,)$ is a probability measure on $(S,\mathscr{S})$;

(b) for each $A \in \mathscr{S}$ $p(\,.\,,A)$ is $\mathscr{S}$-measurable.

Remark 1.2. Let p be a transition probability on a measurable space $(S,\mathscr{S})$.

i) If $f : S \to \mathbb{R}$ is $\mathscr{S}$-$\mathscr{B}^1$-measurable and bounded, then also $g := \int_S f(x)\,p(\,.\,,dx)$ is.

ii) If μ is a probability measure on $(S,\mathscr{S})$, then also $\nu := \int_S p(x,\,.\,)\,\mu(dx)$ is.

Proof. i) By 1.1(b) the statement holds for indicator functions $f = 1_A$ of measurable sets, hence also (linearity of the integral) for step functions.
If $f \geq 0$, there exist approximating step functions $0 \leq f_n \nearrow f$; here $g_n := \int f_n(x)\,p(\,.\,,dx)$ is measurable (since f_n is a step function) and (by the bound of f) bounded; on the other hand (monotone convergence) $g_n \nearrow \int f(x)\,p(\,.\,,dx) \equiv g$, such that g is measurable as pointwise limit of measurable functions; g is bounded with the same bound as f.
More generally, decompose $f = f^+ - f^-$ mit $f^+, f^- \geq 0$; by what we know, $g^\pm := \int f^\pm(x)\,p(\,.\,,dx)$ measurable and bounded, hence also $g = g^+ - g^-$.

ii) For a sequence $(A_n)_{n\in\mathbb{N}}$ of pairwise disjoint $A_n \in \mathscr{S}$ we have

$$\nu\left(\dot{\bigcup_n} A_n\right) \equiv \int_S p\left(x, \dot{\bigcup} A_n\right)\ \mu(dx) \overset{1.1(a)}{=} \int_S \sum_n p(x,A_n)\ \mu(dx)$$
$$\overset{\text{mon. conv.}}{=} \sum_n \int_S p(x,A_n)\ \mu(dx) \equiv \sum_n \nu(A_n)\ \ ;$$

moreover $\nu(S) \equiv \int_S p(x,S)\,\mu(dx) \overset{1.1(a)}{=} \int_S \mu(dx) = 1$ (μ probability measure). □

Definition 1.3. Let S be a Polish space, and $(p_n)_{n\in\mathbb{N}}$ a sequence of transition probabilities, μ a probability measure. Then let $P_0 := \mu$ and

$$P_n(B_0 \times \cdots \times B_n) := \int_{B_0\times\cdots\times B_n} p_n(x_{n-1},dx_n)p_{n-1}(x_{n-2},dx_{n-1})\cdots p_1(x_0,dx_1)\,\mu(dx_0)$$

for $n \in \mathbb{N}$ and $B_i \in \mathscr{S} \equiv \mathscr{B}(S)$.

By remark 1.2 we obtain recursively that P_n is well defined on the semiring[1]

$$\mathscr{R}_n := \{B_0 \times \cdots \times B_n : B_i \in \mathscr{S}\}\,.$$

Moreover, P_n can be extended to a measure on the σ-algebra generated by the ring $r(\mathscr{R}_n)$

$$\sigma\left(r(\mathscr{R}_n)\right) = \underbrace{\mathscr{S} \otimes \cdots \otimes \mathscr{S}}_{(n+1)\text{times}} \equiv \mathscr{S}^{n+1} \quad \left(\overset{S \text{ Polish}}{=} \mathscr{B}(S^{n+1}) \right)\,.$$

PROOF. P_n induces a pre-measure on the ring $r(\mathscr{R}_n)$. By Caratheodory, to extend it to $\sigma\left(r(\mathscr{R}_n)\right)$ we have to prove that P_n is σ-additive on the ring. Since P_n is a finite pre-measure, σ-additivity is equivalent to the property of *continuity from above* that will be shown in what follows; by recursion and remark 1.2 it is enough to consider the case $n = 1$:
Let $(A_k)_{k\in\mathbb{N}}$ be a sequence in $r(\mathscr{R}_1)$ with $A_k \searrow \emptyset$. We have to prove: $P_1(A_k) \xrightarrow{k\to\infty} 0$. Denoting the intersection by $A \in r(\mathscr{R}_1)$ near $x \in S$ with

$$A_x := \{y \in S : (x, y) \in A\}\,,$$

we obtain by $A_k \searrow \emptyset$ for all $x \in S$:

$$(A_k)_x \searrow \emptyset \quad (k \to \infty)\,.$$

By *continuity from above* of the measure $p_1(x, \,.\,)$ we get

$$p_1(x, (A_k)_x) \xrightarrow{k\to\infty} 0 \qquad (x \in S)$$

and thus by dominated convergence:

$$P_1(A_k) = \int_S p_1(x, (A_k)_x)\; \mu(dx) \xrightarrow{k\to\infty} 0\,.$$

□

Our next aim is the construction of a Markov chain on $S^{\mathbb{N}_0}$ with transition probabilities $(p_n)_{n\in\mathbb{N}}$ and initial distribution μ. For this, let S be Polish, endowed with the Borel σ-algebra $\mathscr{B}(S) =: \mathscr{S}$. We verify the consistency condition by Kolmogorov for $(P_n)_{n\in\mathbb{N}_0}$. For this purpose we define:

For $F, G \subset \mathbb{N}_0$ with $F \subset G$ let

$$\begin{aligned} \pi_{G,F} \;:\; S^G &\longrightarrow S^F \\ (x_i)_{i\in G} &\longmapsto (x_i)_{i\in F} \end{aligned}$$

[1] A collection of sets is called $\mathscr{P}$ *semiring*, if (cf. Halmos [HM 74, S.22]):

- for $E \in \mathscr{P}$ and $F \in \mathscr{P}$ also $E \cap F \in \mathscr{P}$, and
- for $E \in \mathscr{P}$ and $F \in \mathscr{P}$ mit $E \subset F$ there exist pairwise disjoint sets $C_1, \ldots, C_n \in \mathscr{P}$, such that

$$F \setminus E = \cup_{i=1}^n C_i.$$

the projection on the smaller index set and herewith $\pi_F := \pi_{\mathbb{N}_0,F}$; correspondingly let $m, n \in \mathbb{N}_0$ with $m \leq n$

$$\begin{array}{rcl} \pi_{n,m} \ : \ S^{n+1} & \longrightarrow & S^{m+1} \\ (x_0, \dots, x_n) & \longmapsto & (x_0, \dots, x_m) \end{array}$$

and for $m \in \mathbb{N}_0$

$$\begin{array}{rcl} \pi_m \ : \ S^{\mathbb{N}_0} & \longrightarrow & S^{m+1} \\ (x_i)_{i\in\mathbb{N}_0} & \longmapsto & (x_0, \dots, x_m)\,. \end{array}$$

These projections are measurable for the respective product σ-algebras. The uniqueness theorem for measures (applied to $\cap$-stable generators of σ-algebras consisting of cylinder sets) yields for $m, n \in \mathbb{N}_0$ with $m \leq n$ the equality

$$P_n \circ \pi_{n,m}^{-1} \ = \ P_m$$

of measures on $\mathscr{S}^{m+1}$; analogously for finite $F, G \subset \mathbb{N}_0$ with $F \subset G$ also

$$P_G \circ \pi_{G,F}^{-1} \ = \ P_{\max F} \circ \pi_{\{0,\dots,\max F\},F}^{-1} \ =: \ P_F \ ;$$

this consistency property implies that $(P_F)_{F\subset\mathbb{N}_0 \text{ finite}}$ defines a pre-measure on $(S^{\mathbb{N}_0}, \mathscr{B}(S)^{\mathbb{N}_0})$. By Kolmogorov's consistency theorem it is even σ-additive. Therefore there is a unique probability measure P_μ on $(S^{\mathbb{N}_0}, \mathscr{B}(S)^{\mathbb{N}_0})$ with

$$P_\mu \circ \pi_n^{-1} \ = \ P_n \qquad (n \in \mathbb{N}_0)\,. \tag{1}$$

Satz 1.4 (canonical Markov chain)**.** *Let S be a Polish space with transition probabilities $(p_n)_{n\in\mathbb{N}}$ and probability measure μ; let P_μ be the thus induced probability measure on $S^{\mathbb{N}_0}$. Then*

$$X_n \ := \ \pi_{\{n\}} \ \equiv \ \pi_{\mathbb{N}_0,\{n\}} \qquad (n \in \mathbb{N}_0)$$

is a Markov chain *on*

$$(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_n)_{n\in\mathbb{N}_0}) \ := \ \left(S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0}, P_\mu, (\sigma(\pi_n))_{n\in\mathbb{N}_0}\right)\,,$$

i.e. we have

i) X_n is $\mathscr{F}_n$-measurable, and

ii) for all $n \in \mathbb{N}_0$ and $B \in \mathscr{S}$ we have:

$$\mathbb{P}\,(X_{n+1} \in B \,|\, \mathscr{F}_n) \ = \ \mathbb{P}\,(X_{n+1} \in B \,|\, X_n) \ = \ p_{n+1}(X_n, B)\,.$$

Proof. i) X_n is measurable with respect to $\sigma(X_n) \subset \sigma(X_0, \ldots, X_n) \equiv \sigma(\pi_n) \equiv \mathscr{F}_n$.
ii) We have to show:

$$\int_A 1_{\{X_{n+1} \in B\}} \, dP_\mu \;=\; \int_A p_{n+1}(X_n, B) \, dP_\mu \qquad (A \in \mathscr{F}_n)$$

(then also $\mathbb{P}(X_{n+1} \in B | X_n) = p_{n+1}(X_n, B)$, by measurability with respect to $\sigma(X_n) \subset \mathscr{F}_n$). Since $\pi_n^{-1}(\mathscr{R}_n)$ is a $\cap$-stable generator of $\mathscr{F}_n$, it is enough, to prove the above equation for

$$A \;=\; \pi_n^{-1}(B_0 \times \cdots \times B_n) \;\equiv\; \{X_0 \in B_0, \ldots, X_n \in B_n\}$$

with $B_0, \ldots, B_n \in \mathscr{S}$, namely:

$$\begin{aligned}
\int_A 1_{\{X_{n+1} \in B\}} \, dP_\mu \quad &= \quad P_\mu\{X_0 \in B_0, \ldots, X_n \in B_n, X_{n+1} \in B\} \\
&\overset{(1)}{=} \quad P_{n+1}(B_0 \times \cdots \times B_n \times B) \\
&\overset{1.3}{=} \quad \int_{B_0 \times \cdots \times B_n} p_{n+1}(x_n, B) \, P_n(dx_0, \ldots, dx_n) \\
&\overset{\text{transf. thm}}{=} \quad \int_A p_{n+1}(X_n, B) \, dP_\mu \, .
\end{aligned}$$

□

Definition 1.5. Let $(S, \mathscr{S})$ be a measurable space. Then on the path space $\Omega \equiv S^{\mathbb{N}_0}$ the family $\theta \equiv (\theta_n)_{n \in \mathbb{N}_0}$ of *(canonical) shifts* $\theta_n : \Omega \longrightarrow \Omega$ $(n \in \mathbb{N}_0)$ is defined by

$$\theta_n(\omega) \;:=\; \big(m \mapsto \omega(m+n)\big) \, .$$

Each θ_n is measurable with respect to $\mathscr{F} \equiv \mathscr{S}^{\mathbb{N}_0}$.

We next prove the Markov property (with deterministic times) and then the strong Markov property (with stopping times). For this we denote $\mathbb{E}_\mu$ resp. $\mathbb{E}_x \equiv \mathbb{E}_{\delta_x}$ the conditional expectations with respect to P_μ resp. P_{δ_x} on Ω with underlying transition probabilities $(p_n)_{n \in \mathbb{N}}$. For simplicity we assume that the Markov chain is time homogeneous:

Definition 1.6. In the situation of Proposition 1.4 the Markov chain is called X *time homogeneous*, if for all $n \in \mathbb{N}$ we have $p_n = p_1 (=: p)$.

Theorem 1.7 (Markov property)**.** *In the situation of 1.4 let the Markov chain X be time homogeneous; let Y be a bounded, $\mathscr{F}$-measurable random variable on Ω. Then*

$$\mathbb{E}_\mu(Y \circ \theta_n \mid \mathscr{F}_n) \;=\; \mathbb{E}_{X_n}(Y) \;\equiv\; \mathbb{E}_x(Y)\Big|_{x = X_n} \qquad (n \in \mathbb{N}_0) \, .$$

Proof. Note first that $\mathbb{E}_{X_n}(Y)$ is indeed measurable with respect to $\mathscr{F}_n$; this follows from adaptedness of X and measurability of $x \mapsto \mathbb{E}_x(Y)$ [this one by definition and recursive application of 1.2 i) clear for indicator functions $Y = 1_{\pi_n^{-1}[B_0 \times \cdots \times B_n]}$ for $B_i \in \mathscr{S}$; the general property follows from the monotone class theorem, since by monotone convergence $\{Y : x \mapsto \mathbb{E}_x(Y) \text{ measurable}\}$ is closed for monotone operations]. So we have to prove the claimed equation. By monotone class arguments, it is enough to argue for the case Y of the form $\prod_{k=0}^{m} g_k(X_k)$ with bounded $\mathscr{S}$-measurable random variables $g_0, \dots, g_m$.

1) We first consider sets $\mathscr{F}_n$ of the form $A := \pi_n^{-1}[A_0 \times \cdots \times A_n]$ with $A_0, \dots, A_n \in \mathscr{S}$; we have

$$\begin{aligned}
\mathbb{E}_\mu(Y \circ \theta_n \cdot 1_A) \quad &\equiv \quad \mathbb{E}_\mu\left(\prod_{k=0}^{m} g_k(X_{n+k}) \cdot 1_A\right) \\
&\overset{(1),1.3}{=} \int_{A_0} \mu(dx_0) \int_{A_1} p(x_0, dx_1) \cdots \int_{A_n} p(x_{n-1}, dx_n) \ \times \\
&\qquad \times \int_S g_0(x_{n+1})\, p(x_n, dx_{n+1}) \cdots \int_S g_m(x_{n+m})\, p(x_{n+m-1}, dx_{n+m}) \\
&\overset{\text{transf. thm}}{=} \mathbb{E}_\mu\left(\mathbb{E}_{X_n}\left(\prod_{k=0}^{m} g_k(X_k)\right) \cdot 1_A\right) \\
&\equiv \quad \mathbb{E}_\mu\big(\mathbb{E}_{X_n}(Y) \cdot 1_A\big)\ ,
\end{aligned}$$

hence the claim follows for all $A \in \mathscr{F}_n$ of the special considered form.

2) Let now $\mathscr{L} := \big\{A \in \mathscr{F}_n : \text{claim from 1) valid for } A\big\}$.According to 1) $\pi_n^{-1}(\mathscr{R}_n) \subset \mathscr{L}$; since $\pi_n^{-1}(\mathscr{R}_n)$ is $\cap$-stable, Dynkin's lemma yields $\mathscr{F}_n = \sigma(\pi_n^{-1}(\mathscr{R}_n)) \subset \mathscr{L}$. □

Our next goal is to extend the Markov property to stopping times.

Definition 1.8. Let $\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\in\mathbb{N}_0})$ be a filtered measure space; $N : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called $(\mathscr{F}_n)_{n\in\mathbb{N}_0}$-*stopping time*, if $\{N \le n\} \in \mathscr{F}_n$ for all $\mathbb{N}_0$. Equivalently, $\{N = n\} \in \mathscr{F}_n$ for all $n \in \mathbb{N}_0$.

To an $(\mathscr{F}_n)_n$-stopping time N we associate the σ-algebra

$$\mathscr{F}_N \ := \ \left\{ A \in \mathscr{F} \ : \ A \cap \left\{ N \overset{(=)}{\le} n \right\} \in \mathscr{F}_n \text{ for all } n \in \mathbb{N}_0 \right\}\ ;$$

it is called N-*past* or σ-*algebra of events before* N.

In the situation from 1.4 and 1.5 we formally enlarge Ω by $\Delta \notin \Omega$, increase $\mathscr{F}$ by $\{\Delta\}$ and define a $(\mathscr{F}_n)_{n\in\mathbb{N}_0}$-stopping time N by

$$\theta_N(\omega) \ := \ \begin{cases} \theta_{N(\omega)}(\omega) & , N(\omega) < \infty \\ \Delta & , N(\omega) = \infty\ . \end{cases}$$

For a random variable Y on Ω let $Y(\Delta) := 0$.

Theorem 1.9 (strong Markov property)**.** *In the situation from 1.4 assume the Markov chain X is time homogeneous; let θ be the shift from 1.5 and N a $(\mathscr{F}_n)_n$-stopping time. If $(Y_n)_{n\in\mathbb{N}_0}$ is a family of $\mathscr{F}$-measurable and (uniformly in (n,ω)) bounded random variables, we have*

$$\mathbb{E}_\mu(Y_N \circ \theta_N \mid \mathscr{F}_N) \ = \ \mathbb{E}_{X_N}(Y_N) \qquad \textit{on } \{N < \infty\}\ ;$$

In particular, for a $\mathscr{F}$-measurable bounded random variable Y :

$$\mathbb{E}_\mu(Y \circ \theta_N \mid \mathscr{F}_N) \;=\; \mathbb{E}_{X_N}(Y) \qquad \text{on } \{N < \infty\}\,.$$

PROOF. Note first that $\omega \mapsto \mathbb{E}_{X_{N(\omega)}(\omega)}(Y_{N(\omega)})$ is in fact $\mathscr{F}_N$-measurable, since it is the composition of the measurable maps $\omega \mapsto (\omega, N(\omega))$, $(\omega, n) \mapsto (X_n(\omega), n)$ and $(x, n) \mapsto \mathbb{E}_x(Y_n)$.
With $A \in \mathscr{F}_N$ we then have

$$\begin{aligned}
\mathbb{E}_\mu\left(Y_N \circ \theta_N \cdot 1_{A\cap\{N<\infty\}}\right) &\overset{\text{dom. conv.}}{=} \sum_{n=0}^{\infty} \mathbb{E}_\mu\left(Y_n \circ \theta_n \cdot 1_{A\cap\{N=n\}}\right) \\
&\overset{\text{MP 1.7}}{=} \sum_{n=0}^{\infty} \mathbb{E}_\mu\left(\mathbb{E}_{X_n}(Y_n) \cdot 1_{A\cap\{N=n\}}\right) \\
&\overset{\text{dom. conv.}}{=} \mathbb{E}_\mu\left(\mathbb{E}_{X_N}(Y_N) \cdot 1_{A\cap\{N<\infty\}}\right)\,.
\end{aligned}$$

□

We next aim at investigating invariant measures of a Markov chain. Invariant measures are strongly correlated with return properties. We therefore assume as a further simplification that S is countable; for the representation with general Polish state space S see Meyn & Tweedie [M-T 93].

Let in the following

$$T_y \;:=\; \inf\{\, n \in \mathbb{N} \;:\; X_n = y \,\} \qquad (y \in S)\,,$$

be the first hitting time of y and thus

$$\rho_{xy} \;:=\; P_x(T_y < \infty) \qquad (x, y \in S)\,.$$

$y \in S$ is called $\left\{\begin{array}{c}\textit{recurrent}\\ \textit{transient}\end{array}\right\}$, if $\left\{\begin{array}{c}\rho_{yy} = 1\\ \rho_{yy} < 1\end{array}\right\}$. The number of visits in y,

$$H_y \;:=\; \sum_{n=1}^{\infty} 1_{\{X_n = y\}}$$

characterizes recurrence and transience of y in the following way:

Theorem 1.10 (transience and recurrence)**.** *Let the Markov chain X from 1.4 be time homogeneous with countable state space S. Then for $y \in S$:*

$$\begin{aligned}
y \ \textit{transient} &\implies \mathbb{E}_x(H_y) \;=\; \frac{\rho_{xy}}{1-\rho_{yy}} \;<\; \infty \quad (\forall\, x \in S)\,,\\
y \ \textit{recurrent} &\iff \mathbb{E}_y(H_y) \;=\; \infty\,.
\end{aligned}$$

PROOF. For $k \in \mathbb{N}$ let T_y^k the time of the k-th visit in y. With this we have

$$P_x(T_y^k < \infty) \; = \; \rho_{xy} \cdot \rho_{yy}^{k-1} \qquad (x \in S, k \in \mathbb{N})\,; \qquad (\star)$$

for $k = 1$ this is just the definition of ρ_{xy}; for $k > 1$ it follows inductively:

$$\begin{aligned} P_x(T_y^k < \infty) &= P_x\left(T_y^{k-1} < \infty\,,\, T_y \circ \theta_{T_y^{k-1}} < \infty\right) \\ &= \mathbb{E}_x\left(1_{\{T_y^{k-1}<\infty\}} \underbrace{\mathbb{E}_x\left(1_{\left\{T_y \circ \theta_{T_y^{k-1}} < \infty\right\}} \,\Big|\, \mathscr{F}_{T_y^{k-1}}\right)}_{\overset{\text{str. MP 1.9}}{=} \mathbb{E}_y\left(1_{\{T_y<\infty\}}\right) \,\equiv\, \rho_{yy}}\right) \\ &= \rho_{yy} \;\cdot\; P_x(T_y^{k-1} < \infty) \\ &\overset{\text{ind. hyp.}}{=} \rho_{xy}\rho_{yy}^{k-1} \;. \end{aligned}$$

Therefore

$$\mathbb{E}_x(H_y) \; = \; \sum_{n=1}^{\infty} P_x \underbrace{\{H_y \geq n\}}_{\{T_y^n<\infty\}} \; \overset{(\star)}{=} \; \rho_{xy} \sum_{n=1}^{\infty} \rho_{yy}^{n-1} \; = \; \frac{\rho_{xy}}{1 - \rho_{yy}} \;;$$

the geometric series converges for $\rho_{yy} < 1$, and diverges iff $\rho_{yy} = 1$. □

We next show that recurrence is contagious:

Theorem 1.11. *Let the Markov chain X from 1.4 be time homogeneous with countable S. If $x \in S$ is recurrent and $\rho_{xy} > 0$ with some $y \in S$, then also y is recurrent and we have $\rho_{yx} = 1$.*

Proof. By recurrence of von x we have

$$\begin{aligned} 0 = P_x(T_x=\infty) &\geq P_x\left(T_y<\infty\,,\, T_x\circ\theta_{T_y}=\infty\right) \\ &= \mathbb{E}_x\Bigg(1_{\{T_y<\infty\}} \underbrace{\mathbb{E}_x\left(1_{\{T_x\circ\theta_{T_y}=\infty\}}\,\Big|\,\mathscr{F}_{T_y}\right)}_{\overset{\text{str. MP 1.9}}{=}\ \mathbb{E}_y\left(1_{\{T_x=\infty\}}\right)\ =\ (1-\rho_{yx})}\Bigg) \\ &= \rho_{xy}\,(1-\rho_{yx})\ ; \end{aligned}$$

Since by hypothesis $\rho_{xy}>0$, we obtain: $\rho_{yx}=1$.
With this the recurrence of y follows: by $\rho_{xy}>0$ and $\rho_{yx}=1$ there exist $k_1,k_2\in\mathbb{N}$ with

$$P_x(X_{k_1}=y)>0 \quad \text{and} \quad P_y(X_{k_2}=x)>0\,.$$

By Chapman-Kolmogorov for $n\in\mathbb{N}$ we have:

$$P_y(X_{n+k_1+k_2}=y)\ \geq\ P_y(X_{k_2}=x)\,P_x(X_n=x)\,P_x(X_{k_1}=y)\,,$$

hence

$$\mathbb{E}_y(H_y)\ =\ \sum_{n=1}^{\infty}P_y(X_n=y)\ \geq\ \underbrace{P_y(X_{k_2}=x)}_{>0}\ \underbrace{\mathbb{E}_x(H_x)}_{\overset{1.10}{=}\infty}\ \underbrace{P_x(X_{k_1}=y)}_{>0}\ .$$

Hence also $\mathbb{E}_y(H_y)=\infty$ and y is recurrent by 1.10. □

Thus the set of recurrent states decomposes into classes: For $x,y\in S$ let

$$x\sim y\ \ :\Longleftrightarrow\ \ \big(x=y\ \text{ or }\ (\rho_{xy}>0 \text{ and } \rho_{yx}>0)\big)\,.$$

Theorem 1.12. *Let the Markov chain X from 1.4 be time homogeneous with countable S. Then the set of recurrent states $R:=\{x\in S:\rho_{xx}=1\}$ decomposes into a family $(R_i)_{i\in I}$ of pairwise disjoint classes, the equivalence classes of $\sim$.*

Proof. We have to show that $\sim$ is an equivalence relation: reflexivity and symmetry follow directly from the definition, so that only transitivity remains to prove:
If $x,y,z\in R$ are fixed, we have to show that with $x\sim y$ and $y\sim z$ also $x\sim z$ holds true. For this purpose we may wlog assume $x\neq y$ and $x\neq z$; by definition of $\sim$ we have $\rho_{xy}>0$ and $\rho_{yz}>0$. Applying the strong Markov property as in the proofs of 1.10 and 1.11 we obtain:

$$\rho_{xz}\ \equiv\ P_x(T_z<\infty)\ \geq\ P_x(T_y<\infty\,,\,T_z\circ\theta_{T_y}<\infty)\ =\ \rho_{xy}\,\rho_{yz}\ >\ 0\,,$$

whence with 1.11 we get $(x\in R)$: $\rho_{zx}=1>0$, in summary $x\sim z$. □

2. Invariant measures and asymptotic behavior

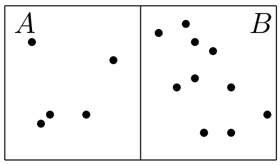
We further consider the following situation: The countable space S is state space of a canonical time homogeneous Markov chain $(X_n)_{n\in\mathbb{N}_0}$ with space of trajectories $(\Omega, \mathscr{F}) := (S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0})$ and transition matrix p.

Definition 2.1. A measure μ on $\mathscr{S}$ is called *stationary*, if for all $y \in S$ we have

$$\mu(y) \;=\; \big(\mu p\big)(y) \;\equiv\; \sum_{x\in S} \mu(x)\, p(x,y) \;<\; \infty\,.$$

A measure μ on $\mathscr{S}$ is called *invariant*, if it is a stationary probability measure.

Example 2.2 (Ehrenfest model of diffusions)**.** In a system consisting of the containers A and B we have a total of r molecules. Let X_n be the number of molecules in A at time $n \in \mathbb{N}_0$. This quantity takes its values in $S := \{0, 1, \dots, r\}$. By



$$p(k,m) := \begin{cases} \frac{r-k}{r} & , \; m = k+1 \\ \frac{k}{r} & , \; m = k-1 \\ 0 & , \; \text{else} \end{cases}$$

a transition probability on S is defined, that is proportional to the number of molecules in container A. For this transition matrix the binomial distribution on S,

$$\mu(k) := \binom{r}{k} 2^{-r} \qquad (k \in S \equiv \{0, 1, \dots, r\})\,,$$

is an invariant measure.

Proof. Since μ is a probability measure, we only have to show that $\mu(k) = \sum_{m=0}^{r} p(m,k)\, \mu(m)$ for $k = 0, 1, \dots, r$ is valid. With $k = 1, \dots, r-1$ we have

$$\begin{aligned} \sum\nolimits_{m=0}^{r} p(m,k)\, \mu(m) \;&=\; p(k+1,k)\, \mu(k+1) \;+\; p(k-1,k)\, \mu(k-1) \\ &\equiv\; 2^{-r} \left[\binom{r}{k+1} \frac{k+1}{r} \;+\; \binom{r}{k-1} \frac{r-(k-1)}{r} \right] \\ &=\; 2^{-r} \left[\frac{(r-1)!}{k!\,(r-(k+1))!} \;+\; \frac{(r-1)!}{(k-1)!\,(r-k)!} \right] \\ &=\; 2^{-r} \frac{(r-1)!}{(k-1)!\,(r-k-1))!} \left[\frac{1}{k} \;+\; \frac{1}{r-k} \right] \\ &=\; 2^{-r} \frac{r!}{k!\,(r-k)!} \\ &\equiv\; \mu(k)\;. \end{aligned}$$

In the cases $k = 0$ and $k = r$ only one summand does not vanish. □

Now we show how to associate with each class of recurrent states a stationary measure; the Markov chain decouples on these classes. We constantly use

$$P_x(X_n = y) \;=\; p^n(x,y) \qquad (x, y \in S\,;\, n \in \mathbb{N})\,,$$

where $p^n(x,y)$ is the n-fold matrix product.

Theorem 2.3. *Let x be recurrent and $T \equiv T_x := \inf\{n \in \mathbb{N} : X_n = x\}$ its first hitting time. Then*

$$\mu(y) := \mathbb{E}_x\left(\sum_{n=0}^{T-1} 1_{\{X_n=y\}}\right) = \sum_{n=0}^{\infty} P_x(X_n = y\,,\, T > n) \qquad (y \in S)$$

defines a stationary measure.

Proof. First we prove the equation $\mu p = \mu$; this way we prove that $\mu(y) < \infty$ for all $y \in S$. Note that $\mu(x) = 1$.

(a) $\sum_{y\in S} \mu(y)\, p(y,z) = \mu(z)$ for all $z \in S$:

1) If $z \neq x$, the Markov property (Thm 1.7) implies:

$$\begin{aligned}
\sum_{y\in S} \mu(y)\, p(y,z) &\overset{\text{Fubini}}{=} \sum_{n=0}^{\infty}\sum_{y\in S} P_x(X_n = y\,,\, T > n)\ \cdot\ P_y(X_1 = z)\\
&\overset{\text{MP}}{=} \sum_{n=0}^{\infty}\sum_{y\in S} P_x(X_n = y\,,\, T > n\,,\, X_{n+1} = z)\\
&= \sum_{n=0}^{\infty} P_x(T > n\,,\, X_{n+1} = z)\\
&\overset{z\neq x}{=} \sum_{n=0}^{\infty} P_x(T > n+1\,,\, X_{n+1} = z)\\
&= \sum_{n=1}^{\infty} P_x(T > n\,,\, X_n = z)\\
&\overset{z\neq x}{=} \sum_{n=0}^{\infty} P_x(T > n\,,\, X_n = z) \qquad \equiv \qquad \mu(z)\,.
\end{aligned}$$

2) If $z = x$, again by the Markov property 1.7:

$$\begin{aligned}
\sum_{y\in S} \mu(y)\, p(y,x) &\overset{\text{ME}}{=} \sum_{n=0}^{\infty}\sum_{y\in S} P_x(X_n = y\,,\, T > n\,,\, X_{n+1} = x)\\
&= \sum_{n=0}^{\infty} P_x(T = n+1) \quad = \quad \rho_{xx} \overset{x\text{ rec.}}{=} 1 \ = \ \mu(x)\,.
\end{aligned}$$

(b) $\mu(y) < \infty$ for all $y \in S$:

1) In case $\rho_{xy} > 0$: By iteration of (a) we get: $\mu = \mu\, p^n$ for $n \in \mathbb{N}$ so

$$1 \ = \ \mu(x) \ \overset{(a)}{=} \ \big(\mu\, p^n\big)(x) \ = \ \sum\nolimits_{y\in S} \mu(y)\, p^n(y,x) \qquad (n \in \mathbb{N})\,.$$

Consequently necessarily $\mu(y) < \infty$, if $p^n(y,x) > 0$ with some $n \in \mathbb{N}$; since $p^n(y,x) = P_y(X_n = x)$, the latter is implied by $\rho_{yx} \equiv P_y(T_x < \infty) > 0$, which in the case considered $\rho_{xy} > 0$ by recurrence of x follows from Thm 1.11 (hence $x \sim y$).

2) If $\rho_{xy} = 0$, the definition of μ gives $\mu(y) = 0\,(< \infty)$. □

Theorem 2.4 (Uniqueness of stationary measures)**.** *Let* $(X_n)_{n\in\mathbb{N}_0}$ *be* irreducible*, i.e. S has only one class of recurrent states. Then the stationary measure* μ *from Theorem 2.3 is unique up to multiplication by constants.*

PROOF. Let $a \in S$ be a recurrent state and μ the stationary measure belonging to a according to 2.3. If ν denotes a further stationary measure, we have to show:

$$\nu(z) = \mu(z) \cdot \nu(a) \qquad (z \in S).$$

By stationarity of ν we obtain iteratively for $z \in S$:

$$\begin{aligned}
\nu(z) &= \sum_{y\in S} \nu(y)\, p(y,z) \\
&= \nu(a)\, p(a,z) + \sum_{y\neq a} \nu(y)\, p(y,z) \\
&= \nu(a)\, p(a,z) + \sum_{y\neq a} \left(\sum_{x\in S} \nu(x)\, p(x,y) \right) p(y,z) \\
&= \nu(a)\, p(a,z) + \sum_{y\neq a} \nu(a)\, p(a,y)\, p(y,z) + \sum_{y\neq a}\sum_{x\neq a} \nu(x)\, p(x,y)\, p(y,z) \\
&= \nu(a)\, P_a(X_1 = z) + \sum_{y\neq a} \nu(a)\, P_a(X_1 \neq a\,,\, X_2 = z) \\
&\qquad\qquad + P_\nu(X_0 \neq a\,,\, X_1 \neq a\,,\, X_2 = z) \\
&= \cdots = \\
&= \nu(a) \sum_{m=1}^{n} P_a(X_k \neq a \text{ for } 1 \leq k < m\,,\, X_m = z) \\
&\qquad\qquad + P_\nu(X_0 \neq a\,,\, X_1 \neq a\,,\, \ldots, X_{n-1} \neq a\,,\, X_n = z) \\
&\geq \nu(a) \cdot \mu(z)
\end{aligned}$$

$(n \to \infty)$ by definition of μ; therefore for $n \in \mathbb{N}$:

$$\nu(a) = \sum_{z\in S} \nu(z)\, p^n(z,a) \geq \nu(a) \sum_{z\in S} \mu(z)\, p^n(z,a) = \nu(a)\, \mu(a) = \nu(a).$$

In the previous estimate $\nu(z) \geq \nu(a)\, \mu(z)$ the inequality "'>"' can only be valid if $p^n(z,a) = 0$ for each $n \in \mathbb{N}$. By irreducibility of for any z there exists $n \in \mathbb{N}$ with $p^n(z,a) > 0$. Therefore $\nu(z) = \nu(a)\, \mu(z)$. □

We give a necessary condition for the normability of stationary measures:

Satz 2.5. *If there exists an invariant measure* μ*, all states* y *with* $\mu(y) > 0$ *are recurrent.*

PROOF. For $n \in \mathbb{N}$ we have by stationarity $\mu = \mu p^n$, hence by Fubini

$$\sum_{n=1}^{\infty} \mu(y) = \sum_{x\in S} \mu(x) \sum_{n=1}^{\infty} p^n(x,y) \overset{1.10}{=} \sum_{x\in S} \mu(x) \frac{\rho_{xy}}{1-\rho_{yy}} \leq \frac{\mu(S)}{1-\rho_{yy}}.$$

By hypothesis $\sum_{n=1}^{\infty} \mu(y) = \infty$ and $\mu(S) = 1 < \infty$, hence $\rho_{yy} = 1$. □

Theorem 2.6. *Let $(X_n)_{n\in\mathbb{N}_0}$ be irreducible and μ an invariant measure. Then*

$$\mu(x) \;=\; \frac{1}{\mathbb{E}_x(T_x)} \qquad (x \in S)\,.$$

PROOF. Note first that all elements of S are recurrent: Each element with positive mass w.r.t. μ is recurrent by 2.5; but since X is irreducible, this recurrence transfers to all other elements.
Consequently for each fixed $x \in S$ by 2.3 there exists a stationary measure μ_0:

$$\mu_0(z) \;\equiv\; \sum\nolimits_{n\in\mathbb{N}_0} P_x(X_n = z\,,\, T_x > n) \qquad \text{and} \qquad \mu_0(x) = 1\,.$$

Consequently by Fubini:

$$\sum_{z\in S}\mu_0(z) \;=\; \sum_{n=0}^{\infty}\sum_{z\in S} P_x(X_n = z\,,\, T_x > n) \;=\; \sum_{n=0}^{\infty} P_x(T_x > n) \;=\; \mathbb{E}_x(T_x)\,.$$

By the uniqueness statement in 2.4 this means for the normed measure μ:

$$\mu(y) \;=\; \frac{\mu_0(y)}{\sum_{z\in S}\mu_0(z)} \;=\; \frac{\mu_0(y)}{\mathbb{E}_x(T_x)} \qquad (y \in S)\,,$$

whence by $y = x$ the claim follows, since $\mu_0(x) = 1$. □

$x \in S$ is called *positively recurrent*, if $\mathbb{E}_x(T_x) < \infty$; in the other case x is called *null recurrent.*

"'Positively recurrent"' is stronger than "'recurrent"'. Positive and null recurrence are properties of classes. In the Ehrenfest model 2.2 every state is positively recurrent.

Corollary 2.7. *Let $(X_n)_{n\in\mathbb{N}_0}$ be irreducible. Then the following statements are equivalent:*

i) There exists an invariant measure;

ii) There exists a positively recurrent state;

iii) All states are positively recurrent.

PROOF. iii) ⇒ ii) trivial.
ii) ⇒ i) Let x be positively recurrent. By 2.3 there exists a stationary measure μ_0 with total mass $\mu_0(S) = \sum_{z\in S}\mu_0(z) = \mathbb{E}_x(T_x)$ (proof of 2.6), which by positive recurrence is finite. The norming factor μ is therefore invariant:

$$\mu(y) \;:=\; \frac{\mu_0(y)}{\mathbb{E}_x(T_x)} \;\equiv\; \frac{1}{\mathbb{E}_x(T_x)}\sum\nolimits_{n\in\mathbb{N}_0} P_x(X_n = y\,,\, T_x > n) \qquad (y \in S)\,.$$

i) ⇒ iii) Let μ be the invariant measure. By irreducibility $\mu(x) > 0$ for all $x \in S$ (every state x is recurrent, so that $\mu_0(x) = 1$ for the stationary measure μ_0 given according to 2.3; by 2.4 we therefore must have $\mu(x) > 0$). From 2.6 we conclude: $\mathbb{E}_x(T_x) = \frac{1}{\mu(x)} < \infty$ for each $x \in S$. □

We now discuss criteria under which p^n converges to the invariant measure.

Example 2.8. On $S := \{1, 2\}$ $p := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ defines a transition matrix. We have

$$p^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad p^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv p \qquad (n \in \mathbb{N}) .$$

In this case no convergence of $p^n(x, y)$ is given.

Periodicity prevents convergence to the invariant measure.

Definition 2.9. For a recurrent $x \in S$ let[2]

$$I_x := \{n \in \mathbb{N}_0 : p^n(x, x) > 0\} .$$

$d_x :=$ gcD(I_x) is called *period* of x.

By the Chapman-Kolmogorov equation I_x is a semigroup.
In the above example 2.8 we have $I_1 = I_2 = \{$ gerade Zahlen$\}$ and $d_1 = d_2 = 2$.

Lemma 2.10. *Let $x, y \in S$ be recurrent with $x \sim y$. Then $d_x = d_y$.*

Proof. We show[3]: $d_y \mid d_x$. Since our arguments are symmetric in x and y, this implies the claim, for by switching the roles of x and y we also have $d_x \mid d_y$. Wlog we may assume $x \neq y$. By the equivalence $x \sim y$ we therefore have $\rho_{xy} > 0$ and $\rho_{yx} > 0$; in particular there exist $m, n \in \mathbb{N}$ with $p^m(x, y) > 0$ and $p^n(y, x) > 0$. By Chapman-Kolmogorov this implies

$$p^{n+m}(y, y) \geq p^n(y, x)\, p^m(x, y) > 0 .$$

Hence by the above definition we obtain $d_y \mid n + m$.
Let now an arbitrary $k \in I_x$ be given. By what has just been proved $d_y \mid n+m$ we only have to show that also $d_y \mid n + m + k$. These two statements imply $d_y \mid k$ and thus the claim. By Chapman-Kolmogorov and $k \in I_x$ we get

$$p^{n+k+m}(y, y) \geq p^n(y, x)\, p^k(x, x)\, p^m(x, y) > 0 ,$$

and thus $d_y \mid n + k + m$. □

Definition 2.11. (a) A state $x \in S$ is called *aperiodic*, if $d_x = 1$ holds.
(b) An irreducible, recurrent Markov chain is called *aperiodic*, if each state is aperiodic.

As indicated in the above example, we shall see that aperiodicity is a criterion for the convergence of the transition probabilities to the invariant measure. The proof of this fact is prepared by the following lemma.

Lemma 2.12. *For aperiodic x there exists $m_0 \in \mathbb{N}$ with $p^m(x, x) > 0$ for all $m \geq m_0$.*

[2]Reminder: $p^n(x, y) \equiv P_x(X_n = y)$ for $x, y \in S$ and $n \in \mathbb{N}_0$.
[3]As usual "'|"' abbreviates "'is a divisor of"'.

PROOF. We first prove that there is $N \in \mathbb{N}$ such that $N, N+1 \in I_x$. For this purpose let $n_0, n_0 + k \in I_x$ be fixed. In case $k = 1$ the proof is finished. In case $k \geq 2$ we choose $n_1 \in I_x$ with $k \nmid n_1$ (since $d_x = 1$). For this we have (division with remainder)

$$n_1 \;=\; m\,k + r_1 \qquad (m \in \mathbb{N}_0\,,\; 0 < r_1 < k)$$

and by the semigroup property of I_x

$$(m+1)(n_0+k) \in I_x \qquad \text{and} \qquad (m+1)n_0 + n_1 \in I_x\,.$$

For these two elements we have:

$$\begin{aligned}\Big|(m+1)(n_0+k) \,-\, \big((m+1)n_0+n_1\big)\Big| \;&=\; |(m+1)k \,-\, n_1| \\ &\equiv\; |(m+1)k \,-\, (m\,k + r_1)| \;=\; k - r_1 \;<\; k\,.\end{aligned}$$

If $k - r_1 = 1$, the claim holds with $N := (m+1)n_0 + n_1$. If $k - r_1 > 1$, we repeat the step performed with $\tilde{n}_0 := (m+1)n_0 + n_1$ and $\tilde{k} := k - r_1$. After finitely many iterations we obtain $N \in \mathbb{N}$ with $N, N+1 \in I_x$.
With this the claim of the Lemma follows with $m_0 := N^2$, since for $m \geq m_0$ we have

$$m - N^2 \;=\; k\,N + r \qquad (k \in \mathbb{N}_0\,,\; 0 \leq r < N)$$

(division with remainder), so that

$$m \;=\; N^2 + k\,N + r \;=\; (N - r + k)N + r(1+N) \;\in\; I_x$$

by the semigroup property of I_x. □

Theorem 2.13 (Invariant measure is limit of transition probabilities)**.** *Let the Markov chain $(X_n)_{n\in\mathbb{N}_0}$ be aperiodic and possess the invariant measure μ. Then*

$$p^n(x,y) \xrightarrow{n\to\infty} \mu(y) \;=\; \frac{1}{\mathbb{E}_y(T_y)} \qquad (x, y \in S)\,.$$

PROOF(COUPLING OF PROCESSES, W. DÖBLIN). On $S^2 \equiv S \times S$ setting

$$q\big((x_1,y_1)\,,\,(x_2,y_2)\big) \;:=\; p(x_1,x_2)\,p(y_1,y_2) \qquad (x_1,x_2,y_1,y_2 \in S)$$

defines a transition probability. Let $(X_n, Y_n)_{n\in\mathbb{N}_0}$ be the canonical Markov chain associated with q, that is the Markov chain with state space S^2 on

$$(\Omega, \mathscr{F}, \mathbb{P}) \;:=\; \Big((S^2)^{\mathbb{N}_0},\, (\mathscr{S}^2)^{\mathbb{N}_0},\, P_\varrho\Big)\;,$$

where P_ϱ is the probability measure related to q and an initial distribution ϱ (on $\mathscr{S}^2 \equiv \mathscr{S} \otimes \mathscr{S}$) according to Kolmogorov.
With 2.12 we now prove irreducibility of $(X_n, Y_n)_{n\in\mathbb{N}_0}$; from this we get that this coupled process hits the diagonal of S^2 in finite time. This will imply convergence.

1) $(X_n, Y_n)_{n\in\mathbb{N}_0}$ is irreducible: If $x_1, x_2, y_1, y_2 \in S$ are fixed, irreducibility of X provides times $k, l \in \mathbb{N}$ with

$$p^k(x_1, x_2) \;>\; 0 \qquad \text{and} \qquad p^l(y_1, y_2) \;>\; 0\,.$$

Aperiodicity yields according to 2.12 also an $m_0 \in \mathbb{N}$, such that for $m \geq m_0$ we have

$$p^{m+l}(x_2, x_2) \;>\; 0 \qquad \text{and} \qquad p^{m+k}(y_2, y_2) \;>\; 0\,.$$

Hence by Chapman-Kolmogorov also

$$\begin{aligned} q^{k+l+m}\big((x_1, y_1)\,,\,(x_2, y_2)\big)& \\ &\equiv\; p^{k+l+m}(x_1, x_2)\, p^{k+l+m}(y_1, y_2) \\ &\geq\; p^k(x_1, x_2)\, p^{m+l}(x_2, x_2)\;\, p^l(y_1, y_2)\, p^{m+k}(y_2, y_2) \;\;>\;\; 0\,. \end{aligned}$$

Therefore S^2 consists of a unique equivalence class. For irreducibility we have to show that all states in S^2 are recurrent. By 2.5 for this we need a q-invariant measure ν with $\nu(x, y) > 0$ for all $(x, y) \in S^2$. But setting

$$\nu(x, y) \;:=\; \mu(x)\, \mu(y) \qquad\qquad (x, y \in S)$$

provides a q-invariant measure on S^2 by p-invariance of μ:

$$\begin{aligned} \sum_{(x_1,x_2)\in S^2} \nu(x_1, x_2)\, q((x_1, x_2), (y_1, y_2)) \;&\equiv\; \sum_{(x_1,x_2)} \mu(x_1)\, \mu(x_2)\, p(x_1, y_1)\, p(x_2, y_2) \\ =\; \sum_{x_1} \mu(x_1)\, p(x_1, y_1) \sum_{x_2} \mu(x_2)\, p(x_2, y_2) \;&=\; \mu(y_1)\, \mu(y_2) \;\equiv\; \nu(y_1, y_2) \end{aligned}$$

for $(y_1, y_2) \in S^2$; moreover $\nu(y_1, y_2) \;\equiv\; \mu(y_1)\, \mu(y_2) \overset{2.6}{=} \frac{1}{\mathbb{E}_{y_1}(T_{y_1})}\, \frac{1}{\mathbb{E}_{y_2}(T_{y_2})} \overset{2.7\text{ iii)}}{>} 0\,.$

2) Denote by T the first hitting time of the diagonal $D := \{(x, x) : x \in S\}$,

$$T \;:=\; \inf\{\, n \in \mathbb{N} : (X_n, Y_n) \in D\,\}\,,$$

$T_{(x,x)}$ the time of first visit in $(x, x) \in D$. Then on the one hand $T \leq T_{(x,x)}$. If ϱ is an arbitrary initial distribution on S^2, on the other hand by the recurrence proved in 1) we get $T_{(x,x)} < \infty$ P_ϱ- a.s.; in particular $T < \infty$ P_ϱ − a.s. .

X_n and Y_n possess on $\{T \leq n\}$ identical laws ($n \in \mathbb{N}$), since for $y \in S$:

$$\begin{aligned} P_\varrho\,(X_n = y\,,\, T \leq n) \;&=\; \sum_{m=1}^{n} P_\varrho\,(T = m\,,\, X_n = y) \\ &=\; \sum_{m=1}^{n} \sum_{x\in S} P_\varrho\,(T = m\,,\, X_m = x\,,\, X_n = y) \;\;= \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=1}^{n}\sum_{x\in S} P_\varrho\left(X_n = y \,\middle|\, T = m\,,\, X_m = x\right)\, P_\varrho\left(T = m\,,\, X_m = x\right) \\
&\overset{\text{ME}}{=} \sum_{m=1}^{n}\sum_{x\in S} P_\varrho\left(X_n = y \,\middle|\, X_m = x\right)\, P_\varrho\left(T = m\,,\, X_m = x\right) \\
&= \sum_{m=1}^{n}\sum_{x\in S} P_\varrho\left(Y_n = y \,\middle|\, Y_m = x\right)\, P_\varrho\left(T = m\,,\, Y_m = x\right) \\
&= \ldots\ \ldots\ \ldots \overset{\text{same arg.}}{=} P_\varrho\left(Y_n = y\,,\, T \le n\right)\,.
\end{aligned}$$

Here we used that X and Y possess identical transition probability p.

3) Now we prove the claim of the theorem; for this purpose we show the following (stronger) convergence:

$$\sum_{y\in S} |\, p^n(x,y) \;-\; \mu(y)\, | \xrightarrow{n\to\infty} 0$$

for all $x \in S$; the equality $\mu(y) \;=\; 1/\,\mathbb{E}_y(T_y)$ is already clear by 2.6.

For an arbitrary $x \in S$ we fix the initial distribution

$$\varrho \;:=\; \delta_x \otimes \mu$$

on S^2 for the coupled process. Thus for all $y \in S$

$$\begin{aligned}
p^n(x,y) &= P_\varrho(X_n = y) \\
&= P_\varrho\left(X_n = y\,,\, T \le n\right) + P_\varrho\left(X_n = y\,,\, T > n\right) \\
&\overset{2)}{=} P_\varrho\left(Y_n = y\,,\, T \le n\right) + P_\varrho\left(X_n = y\,,\, T > n\right)
\end{aligned}$$

by equality of the laws proven in 2), and

$$\mu(y) \;=\; P_\varrho(Y_n = y) \;\equiv\; P_\varrho\left(Y_n = y\,,\, T \le n\right) + P_\varrho\left(Y_n = y\,,\, T > n\right)$$

by the p-invariance of μ; in summary

$$\begin{aligned}
\sum_{y\in S} |\, p^n(x,y) \;-\; \mu(y)\, | &= \sum_{y\in S} |\, P_\varrho(X_n = y) \;-\; P_\varrho(Y_n = y)\, | \\
&= \sum_{y\in S} |\, P_\varrho\left(X_n = y\,,\, T > n\right) \;-\; P_\varrho\left(Y_n = y\,,\, T > n\right) | \\
&\le \sum_{y\in S} \left[\, P_\varrho\left(X_n = y\,,\, T > n\right) + P_\varrho\left(Y_n = y\,,\, T > n\right) \right] \\
&= 2\, P_\varrho\left(T > n\right) \quad \xrightarrow{n\to\infty} \quad 0\,,
\end{aligned}$$

since T is P_ϱ-a.s. finite, as seen in 2). □

3. Stationary Processes

In this chapter we consider stochastic processes $X = (X_n)_{n\in\mathbb{N}_0}$ on a fixed probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with values in a Polish space S (equipped with the Borel σ-algebra $\mathscr{S} := \mathscr{B}(S)$). This family of $\mathscr{F}$-$\mathscr{S}$-measurable maps can also be considered as a random sequence

$$X : \Omega \longrightarrow S^{\mathbb{N}_0}, \qquad \omega \mapsto (X_n(\omega))_{n\in\mathbb{N}_0},$$

which is $\mathscr{F}$-$\mathscr{S}^{\mathbb{N}_0}$-measurable, with product σ-algebra

$$\mathscr{S}^{\mathbb{N}_0} := \sigma\Big(\bigcup\nolimits_{n\in\mathbb{N}_0} \pi_{\{n\}}^{-1}[B_n] \,:\, B_n \in \mathscr{S}\Big) = \sigma\Big(\bigcup\nolimits_{n\in\mathbb{N}_0} \pi_n^{-1}[B] \,:\, B \in \mathscr{S}^{n+1}\Big);$$

where the second generating system is $\cap$-stable, in contrast to the first. The measure defined by

$$P_X \equiv P_{(X_n)_{n\in\mathbb{N}_0}} := \mathbb{P} \circ X^{-1}$$

on $\mathscr{S}^{\mathbb{N}_0}$ is the *law of* X.
If only distribution properties are relevant, instead of X we can wlog also study its *canonical representation* $(Y)_n := \big(\pi_{\{n\}}\big)_n$ on $(S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0}, P_X)$.

Definition 3.1. A stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ is called *stationary*, if we have:

$$P_{(X_n)_{n\in\mathbb{N}_0}} = P_{(X_{n+k})_{n\in\mathbb{N}_0}} \qquad (\forall\, k \in \mathbb{N}).$$

The distribution of a stationary process does not "'move"'; this will be enforced in the following Lemma:

Lemma 3.2. $X = (X_n)_{n\in\mathbb{N}_0}$ *is stationary iff we have:*

$$P_{(X_0,\dots,X_n)} = P_{(X_k,\dots,X_{k+n})} \qquad (k \in \mathbb{N},\, n \in \mathbb{N}_0).$$

Proof. "'$\Rightarrow$"' For all $k \in \mathbb{N}$, $n \in \mathbb{N}_0$ and $B \in \mathscr{S}^{n+1}$ we have:

$$\begin{aligned} P_{(X_0,\dots,X_n)}(B) &\equiv \mathbb{P}\{(X_0,\dots,X_n) \in B\} \\ &= \mathbb{P}\{(X_m)_{m\in\mathbb{N}_0} \in \pi_n^{-1}(B)\} \\ &\overset{\text{stat}}{=} \mathbb{P}\{(X_{m+k})_{m\in\mathbb{N}_0} \in \pi_n^{-1}(B)\} \\ &= \mathbb{P}\{(X_k,\dots,X_{n+k}) \in B\} \equiv P_{(X_k,\dots,X_{k+n})}(B). \end{aligned}$$

"'$\Leftarrow$"' By hypothesis we have for all $k \in \mathbb{N}$, $n \in \mathbb{N}_0$ and $B \in \mathscr{S}^{n+1}$:

$$P_{(X_m)_{m\in\mathbb{N}_0}}\big(\pi_n^{-1}(B)\big) = P_{(X_{m+k})_{m\in\mathbb{N}_0}}\big(\pi_n^{-1}(B)\big)$$

(see calculation above). But since $\{\bigcup_{n\in\mathbb{N}_0} \pi_n^{-1}(B) : B \in \mathscr{S}^{n+1}\}$ is a $\cap$-stable generator of $\mathscr{S}^{\mathbb{N}_0}$, this implies $P_{(X_m)_{m\in\mathbb{N}_0}} = P_{(X_{m+k})_{m\in\mathbb{N}_0}}$ by the uniqueness theorem for measures. □

Example 3.3 (Markov chain with transition probability p)**.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a Markov chain on a countable space S (equipped with $\mathscr{S} := \mathscr{B}(S) \equiv \mathfrak{P}(S)$) with transition probability p and invariant measure μ. Then $(X_n)_{n\in\mathbb{N}_0}$ is stationary on $(\Omega, \mathscr{F}, \mathbb{P}) := (S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0}, P_\mu)$.

Proof. Note first that we have for all $B := B_0 \times B_1 \times \cdots \times B_n \in \mathscr{S}^{n+1}$:

$$\begin{aligned}
(P_\mu)_{(X_1,\dots,X_{n+1})}(B) &\equiv P_\mu\{ X_1 \in B_0 , X_2 \in B_1 , \dots , X_{n+1} \in B_n \} \\
&= P_\mu\{ X_0 \in S , X_1 \in B_0 , X_2 \in B_1 , \dots , X_{n+1} \in B_n \} \\
&= \sum_{z\in S} \mu(z) \sum_{x_0\in B_0} p(z,x_0) \sum_{x_1\in B_1} p(x_0,x_1) \cdots \sum_{x_n\in B_n} p(x_{n-1},x_n) \\
&= \sum_{x_0\in B_0} \sum_{z\in S} \mu(z)\, p(z,x_0) \sum_{x_1\in B_1} p(x_0,x_1) \cdots \sum_{x_n\in B_n} p(x_{n-1},x_n) \\
&\overset{\text{inv}}{=} \sum_{x_0\in B_0} \mu(x_0) \sum_{x_1\in B_1} p(x_0,x_1) \cdots \sum_{x_n\in B_n} p(x_{n-1},x_n) \\
&= P_\mu\{ X_0 \in B_0 , X_1 \in B_1 , \dots , X_n \in B_n \} \\
&= (P_\mu)_{(X_0,X_1,\dots,X_n)}(B) .
\end{aligned}$$

By k-fold iteration of this argument we obtain the criterion for stationarity from 3.2. □

Example 3.4 (Rotation on circle)**.** Let $(\Omega, \mathscr{F}, \mathbb{P}) := ([0,1), \mathscr{B}[0,1), \lambda\big|_{\mathscr{F}})$, where λ denotes the Lebesgue measure. Then for each fixed $\theta \in [0,1)$ the process $(X_n)_{n\in\mathbb{N}_0}$,

$$X_n : \Omega \longrightarrow S := \Omega , \qquad X_n(\omega) := \omega + n\cdot\theta \pmod 1 , \qquad n \in \mathbb{N}_0 ,$$

is a stationary Markov chain on $(S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0}, P_\lambda)$ with respect to the transition probability

$$p : S \times \mathscr{S} \longrightarrow [0,1] , \qquad p(x,A) := \begin{cases} 1 , & \text{if } y = x + \theta \pmod 1 \in A, \\ 0 , & \text{else.} \end{cases}$$

Proof. By translation invariance of the Lebesgue measure λ is p-invariant, since for $A \in \mathscr{S}$

$$\int_0^1 \lambda(dz)\, p(z,A) = \lambda(A - \theta\,(\text{mod}1)) = \lambda(A).$$

Hence as in Example 3.3 for all $B := B_0 \times B_1 \times \cdots \times B_n \in \mathscr{S}^{n+1}$:

$$\begin{aligned}
(P_\lambda)_{(X_1,\dots,X_{n+1})}(B) &= P_\lambda\{ X_0 \in S , X_1 \in B_0 , X_2 \in B_1 , \dots , X_{n+1} \in B_n \} \\
&= \int_\Omega \lambda(dz) \int_{B_0} p(z,dx_0) \int_{B_1} p(x_0,dx_1) \cdots \int_{B_n} p(x_{n-1},dx_n) \\
&= \int_{B_0} \int_\Omega \lambda(dz)\, p(z,dx_0) \int_{B_1} p(x_0,dx_1) \cdots \int_{B_n} p(x_{n-1},dx_n) \\
&\overset{\text{inv}}{=} \int_{B_0} \lambda(dx_0) \int_{B_1} p(x_0,dx_1) \cdots \int_{B_n} p(x_{n-1},dx_n) \\
&= (P_\lambda)_{(X_0,X_1,\dots,X_n)}(B) ,
\end{aligned}$$

and thus stationarity again from Lemma 3.2. □

Theorem 3.5. *Let the process $(X_n)_{n\in\mathbb{N}_0}$ with Polish state space $(S,\mathscr{S})$ be stationary and let $g : S^{\mathbb{N}_0} \longrightarrow S'$ be $\mathscr{S}^{\mathbb{N}_0}$-$\mathscr{S}'$-measurable, where $(S',\mathscr{S}')$ is also Polish. Then*

$$Y_k \;:=\; g\,(X_k\,,\,X_{k+1}\,,\,\ldots) \qquad (k\in\mathbb{N}_0)$$

is stationary (in S').

Proof. By measurability of g for each $k\in\mathbb{N}_0$ also

$$g_k\;:\;S^{\mathbb{N}_0}\longrightarrow S'\,,\qquad x\mapsto g\circ\theta_k(x),$$

is measurable, where $\theta\equiv(\theta_k)_{k\in\mathbb{N}_0}$ (see 1.5) denotes the measurable shift

$$\theta_k\;:\;S^{\mathbb{N}_0}\longrightarrow S^{\mathbb{N}_0}\,,\qquad (x_n)_n\mapsto(x_{n+k})_n.$$

Let now $B\in(\mathscr{S}')^{\mathbb{N}_0}$ be fixed; by measurability of all g_k also $A := (g_0,g_1,\ldots)^{-1}(B)$ is measurable and by $Y_k = g_k\big((X_n)_n\big)$ we obtain for $m\in\mathbb{N}$:

$$\begin{aligned}
\mathbb{P}_{(Y_k)_{k\in\mathbb{N}_0}}(B) \;&\equiv\; \mathbb{P}\big((Y_k)_{k\in\mathbb{N}_0}\in B\big) \;=\; \mathbb{P}\big((X_n)_{n\in\mathbb{N}_0}\in A\big)\\
&\overset{X\text{ stat}}{=}\; \mathbb{P}\big((X_{n+m})_{n\in\mathbb{N}_0}\in A\big) \;=\; \mathbb{P}\big((Y_{k+m})_{k\in\mathbb{N}_0}\in B\big)\\
&\equiv\; \mathbb{P}_{(Y_{k+m})_{k\in\mathbb{N}_0}}(B)\,,
\end{aligned}$$

hence the stationarity of Y. □

Example 3.6 (Bernoulli-Shift)**.** On $(\Omega,\mathscr{F},\mathbb{P}) := ([0,1),\mathscr{B}[0,1),\lambda\big|_{\mathscr{F}})$ $(Y_n)_{n\in\mathbb{N}_0}$,

$$Y_n\;:\;\Omega\longrightarrow\Omega\,,\qquad Y_n\;:=\;\begin{cases}\mathrm{id}_\Omega\,, & n=0,\\ 2Y_{n-1}\ (\mathrm{mod}\ 1)\,, & n\in\mathbb{N}\,,\end{cases}$$

is stationary.

Proof. Let $(X_n)_{n\in\mathbb{N}_0}$ be a Bernoulli sequence with rate $\frac{1}{2}$, realized as product measure $\tilde{\mathbb{P}}$ on $\tilde{\Omega} := \{0,1\}^{\mathbb{N}_0}$; hence $(X_n)_n$ is a sequence of i.i.d. random variables in $S := \{0,1\}$ with $\tilde{\mathbb{P}}\{X_n = 0\} = \tilde{\mathbb{P}}\{X_n = 1\} = \frac{1}{2}$. Then $(X_n)_n$ is stationary. Moreover

$$g \,:\, \tilde{\Omega} \equiv \{0,1\}^{\mathbb{N}_0} \longrightarrow \Omega \equiv [0,1)\,, \qquad (x_n)_n \mapsto \sum\nolimits_{n=0}^{\infty} x_n\, 2^{-n-1} \pmod 1$$

is measurable and so $\tilde{\mathbb{P}} \circ g^{-1} = \mathbb{P}$ (dyadic intervals may be written as sets of the form $\{X_0 = i_0\,, \ldots\,, X_k = i_k\}$ with $i_0, \ldots, i_k \in \{0,1\}$). Because of Theorem 3.5 we have that

$$Z_k \;:=\; g(X_k\,,\, X_{k+1}\,, \ldots) \qquad (k \in \mathbb{N}_0)$$

is stationary; on the other hand we have:

$$\begin{aligned} 2\,Z_0 \;\equiv\; 2\,g(X_0\,,\, X_1\,, \ldots) \;&=\; X_0 \,+\, \sum\nolimits_{n=1}^{\infty} X_n\, 2^{-n} \pmod 1 \\ &=\; X_0 \,+\, \sum\nolimits_{n=0}^{\infty} X_{n+1}\, 2^{-(n+1)} \pmod 1 \\ &=\; g(X_1\,,\, X_2\,, \ldots) \;\equiv\; Z_1\,; \end{aligned}$$

by iteration we obtain: $2\,Z_{n-1} = Z_n\,(n \in \mathbb{N})$, hence with Z also Y is stationary. □

Definition 3.7 (measure preserving map). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A $\mathscr{F}$-$\mathscr{F}$-measurable mapping $\varphi : \Omega \to \Omega$ is called *measure preserving*, if we have: $\mathbb{P} \circ \varphi^{-1} = \mathbb{P}$.

Remark 3.8. Let φ be measure preserving on $(\Omega, \mathscr{F}, \mathbb{P})$ and $X : \Omega \to S$ a $\mathscr{F}$-$\mathscr{S}$-measurable map with values in a Polish space $(S, \mathscr{S})$. Then $(X_n)_{n\in\mathbb{N}_0}$ with

$$X_n \;:=\; \begin{cases} X\,, & n = 0 \\ X \circ \varphi^n\,, & n \in \mathbb{N} \end{cases}$$

is stationary.

Proof. For $B \in \mathscr{S}^{n+1}$ we have:

$$\begin{aligned} \mathbb{P}_{(X_0,\ldots,X_n)}(B) \;&\equiv\; \mathbb{P}\big((X_0, \ldots, X_n) \in B\big) \\ &\overset{\varphi \text{ m.p.}}{=}\; \mathbb{P}\big((X_0, \ldots, X_n) \circ \varphi^k \in B\big) \;=\; \mathbb{P}_{(X_k,\ldots,X_{k+n})}(B)\,, \end{aligned}$$

so that stationarity follows from Lemma 3.2. □

The situation of the preceding remark does not only provide an example for a stationary sequence. It already depicts the general situation.

Satz 3.9 (standard model for stationary sequences).
Let $(Y_n)_{n\in\mathbb{N}_0}$ be stationary on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in a Polish space $(S, \mathscr{S})$. Then there exists a probability space $(\Omega', \mathscr{F}', \mathbb{P}')$ with a measure preserving map $\varphi : \Omega' \to \Omega'$ and a random variable $X_0 : \Omega' \to S$ such that with $X_n := X_0 \circ \varphi^n$ $(n \in \mathbb{N})$ we have:

$$\mathbb{P}'_{(X_n)_{n\in\mathbb{N}_0}} \;=\; \mathbb{P}_{(Y_n)_{n\in\mathbb{N}_0}}\,.$$

Proof. Let $(\Omega', \mathscr{F}', \mathbb{P}') := (S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0}, P_{(Y_n)_{n\in\mathbb{N}_0}})$ and $X_0 := \pi_{\{0\}}$ (projection at time 0) and $\varphi := \theta_1$ (shift). By stationarity of Y φ is measure preserving, since for $A' \in \mathscr{F}'$ we have:

$$\begin{aligned}\mathbb{P}'\left(\varphi^{-1}(A')\right) &= \mathbb{P}\big((Y_n)_n \in \varphi^{-1}(A')\big)\\ &= \mathbb{P}\big((Y_{n+1})_n \in A'\big)\\ &\overset{Y\text{ stat}}{=} \mathbb{P}\big((Y_n)_n \in A'\big) = \mathbb{P}'(A').\end{aligned}$$

The claimed equality of laws follows from the definition of $\mathbb{P}'$. □

Definition 3.10 (invariant, ergodic)**.** Let φ be a measure preserving mapping on $(\Omega, \mathscr{F}, \mathbb{P})$. $A \in \mathscr{F}$ is called $\left\{\begin{array}{l}\textit{invariant}\\ \textit{strictly invariant}\end{array}\right\}$, if $\left\{\begin{array}{l}\varphi^{-1}(A) = A\ \ \mathbb{P}\text{-a.s.}\\ \varphi^{-1}(A) = A\end{array}\right\}$.
φ is called *ergodic*, if for all $A \in \mathscr{I} := \{$ invariant sets$\}$ we have: $\mathbb{P}(A) \in \{0, 1\}$.

Remark 3.11. i) $\mathscr{I}$ is a σ-algebra (sub-σ-algebra of $\mathscr{F}$);

ii) For $A \in \mathscr{I}$ there exists a strictly invariant set $B \in \mathscr{F}$ with $B = A$ $\mathbb{P}$-a.s. (for example $B := \liminf_{n\to\infty} \varphi^{-n}(A)$).

iii) For $A \in \mathscr{I}$ there exists $B \in \mathscr{T} := \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$ with[4] $B = A$ $\mathbb{P}$-a.s. (for example again $B := \liminf_{n\to\infty} \varphi^{-n}(A)$, since $B = \varphi^{-k}(B) \in \sigma(X_k, X_{k+1}, \dots)$).

[4] $\mathscr{T}$ is the σ-algebra of *terminal events*;

Example 3.12. Let $(X_n)_{n\in\mathbb{N}_0}$ be independent random elements in a Polish space S (wlog defined on the sequence space), i.e.:

$$\mathbb{P} \equiv \mathbb{P}_X \;=\; \bigotimes_{n\in\mathbb{N}_0} \mathbb{P}_{X_n}\;.$$

Then we have $\mathbb{P}(A) \in \{0,1\}$ for $A \in \mathscr{T}$; i.e. the shift $\varphi := \theta_1$ is ergodic.

Example 3.13 (Rotation on the circle)**.** As in 3.4 we consider the transformation

$$\varphi \,:\, \Omega \longrightarrow \Omega\,, \qquad \varphi(\omega) \;:=\; \omega + \theta \;\;(\mathrm{mod}\;1)\,,$$

on the probability space $(\Omega, \mathscr{F}, \mathbb{P}) := ([0,1), \mathscr{B}[0,1), \lambda\big|_{\mathscr{F}})$, where λ denotes the Lebesgue measure. Then φ is ergodic iff θ is irrational.

> Proof. "'$\Rightarrow$"' Let θ be rational, hence $\theta = \frac{m}{n}$ with integers $n \geq m \geq 1$. Moreover let $B \in \mathscr{F} \equiv \mathscr{B}[0,1)$ with $0 < \lambda(B) < \frac{1}{n}$. Then $A := \bigcup_{k=1}^{m-1}(B + \frac{k}{n})$ is invariant, but $0 < \lambda(A) < 1$.
> "'$\Leftarrow$"' This can be proven with a Fourier series argument; see e.g. Shiryaev [Sh 95, p.408] or Kallenberg [KB 97, p.174/9]. □

Example 3.14. Let $(X_n)_{n\in\mathbb{N}_0}$ be the canonical Markov chain on $S := \{1,2,3,4\}$ with transition probability

$$p \;:=\; \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

(p is a stochastic matrix, since the sums over the lines equal 1). A measure μ on S is invariant, if we have:

$$\mu(j) \;=\; \sum_{i=1}^{4} p(i,j)\mu(i) \qquad (j = 1,2,3,4)\,.$$

This is satisfied for instance by the measures

$$\mu_0(1) \;=\; \mu_0(2) \;:=\; \frac{1}{2}\,,\;\; \mu_0(3) \;=\; \mu_0(4) \;:=\; 0$$

and

$$\mu_1(1) \;=\; \mu_1(2) \;:=\; 0\,,\;\; \mu_1(3) \;:=\; \frac{1}{3}\,,\;\; \mu_1(4) \;:=\; \frac{2}{3}\,.$$

But then also each

$$\mu_\beta \;:=\; (1-\beta)\mu_0 \,+\, \beta\mu_1 \qquad (0 \leq \beta \leq 1)$$

is invariant. With respect to the canonical shift $\varphi := \theta_1$ we have now:

$$A \;:=\; \{X_n \in \{1,2\}, n \in \mathbb{N}_0\} \;\in\; \mathscr{I} \qquad \text{and} \qquad B \;:=\; \{X_n \in \{3,4\}, n \in \mathbb{N}_0\} \;\in\; \mathscr{I}\,.$$

Hence we further have: $P_{\mu_\beta}(A) = 1-\beta$ and $P_{\mu_\beta}(B) = \beta$. Consequently φ is ergodic iff $\beta \in \{0,1\}$.

Theorem 3.15. *Let* $(X_n)_{n\in\mathbb{N}_0}$ *with Polish state space* $(S,\mathscr{S})$ *be ergodic and* $g : S^{\mathbb{N}_0} \longrightarrow S'$ *be* $\mathscr{S}^{\mathbb{N}_0}$*-*$\mathscr{S}'$*-measurable, where* $(S',\mathscr{S}')$ *is equally Polish. Then*

$$Y_k \;:=\; g\,(X_k\,,\, X_{k+1}\,,\,\dots) \qquad (k\in\mathbb{N}_0)$$

is ergodic (in S'*).*

Proof. Wlog let again $(\Omega,\mathscr{F},\mathbb{P}) = (S^{\mathbb{N}_0},\mathscr{S}^{\mathbb{N}_0},\mathbb{P}_{(X_n)_{n\in\mathbb{N}_0}})$ and $X_n = \pi_{\{n\}}$ (projection at time n) and $\varphi = \theta_1$ (shift). Equally let $(\Omega',\mathscr{F}',\mathbb{P}') = ((S')^{\mathbb{N}_0},(\mathscr{S}')^{\mathbb{N}_0},\mathbb{P}_{(Y_n)_{n\in\mathbb{N}_0}})$ and $\varphi' = \theta_1$. Moreover denote by $\mathscr{I}$ resp. $\mathscr{I}'$ the systems of invariant sets associated with φ resp. φ'.
Let now $A\in\mathscr{I}'$ be fixed; for $B := (g_0,g_1,\dots)^{-1}(A)$ we then have:

$$\begin{aligned}\varphi^{-1}(B) &= (g_1,g_2,\dots)^{-1}(A)\\ &= (g_0,g_1,\dots)^{-1}\left((\varphi')^{-1}(A)\right)\\ &= (g_0,g_1,\dots)^{-1}(A) \quad\equiv\quad B\,,\end{aligned}$$

hence $B\in\mathscr{I}$. By ergodicity of φ we get: $\mathbb{P}'(A)\equiv\mathbb{P}(B)\in\{0,1\}$. □

Example 3.16 (Bernoulli shift)**.** As in 3.6 we consider i.i.d. random variables $(X_n)_n$ in $S := \{0,1\}$ with $\mathbb{P}\{X_n = 0\} = \mathbb{P}\{X_n = 1\} = \frac{1}{2}$. Moreover let

$$g \;:\; \{0,1\}^{\mathbb{N}_0} \longrightarrow [0,1)\,, \qquad (x_n)_n \mapsto \sum\nolimits_{n=0}^{\infty} x_n\,2^{-n-1}\ (\mathrm{mod}\ 1)\,.$$

By example 3.12 X is ergodic, hence according to Theorem 3.15 also

$$Y_k \;:=\; g\,(X_k\,,\, X_{k+1}\,,\,\dots) \qquad (k\in\mathbb{N}_0)\,.$$

4. Birkhoff's ergodic theorem

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with measure preserving mapping $\varphi : \Omega \to \Omega$ and a random variable $X : \Omega \to \mathbb{R}$. We now study the asymptotic behavior of the stochastic process defined by

$$X_k := X \circ \varphi^k \qquad (k \in \mathbb{N}_0).$$

Theorem 4.1 (Ergodic theorem, Birkhoff)**.** *Let $X \in L^1(\mathbb{P})$. Then we have $\mathbb{P}$-a.s. and in $L^1(\mathbb{P})$:*

$$\frac{1}{n} \sum_{k=0}^{n-1} X \circ \varphi^k \xrightarrow{n\to\infty} \mathbb{E}(X|\mathscr{I}) .$$

The proof is based on the following estimate:

Lemma 4.2 (Maximal-ergodic Lemma, Hopf)**.** *In the situation from 4.1 let*

$$\begin{aligned} S_n &:= X_0 + \dots + X_{n-1} \equiv \sum_{k=0}^{n-1} X \circ \varphi^k && (n \in \mathbb{N}) \text{ and} \\ M_n &:= \max\{0, S_1, \dots, S_n\} && (n \in \mathbb{N}_0) . \end{aligned}$$

Then we have:

$$\mathbb{E}\left(X\, 1_{\{M_n > 0\}} \right) \geq 0 \qquad (n \in \mathbb{N}_0) .$$

Proof. In case $n = 0$ nothing is to be proven.
1. First we show:

$$X\, 1_{\{M_n>0\}} \geq 1_{\{M_n>0\}} (M_n - M_n \circ \varphi) \qquad (n \in \mathbb{N}) ;$$

by the definitions above we have $S_k - M_n \leq 0$ for all $k \in \{1, \dots, n\}$, hence also

$$X \geq X + (S_k - M_n) \circ \varphi = (X + S_k \circ \varphi) - M_n \circ \varphi \equiv S_{k+1} - M_n \circ \varphi$$

and thus

$$X \geq \max\{S_1, \dots, S_n\} - M_n \circ \varphi ;$$

in particular we have shown:

$$\begin{aligned} X\, 1_{\{M_n>0\}} &\geq 1_{\{M_n>0\}} \max\{S_1, \dots, S_n\} - 1_{\{M_n>0\}} M_n \circ \varphi \\ &= 1_{\{M_n>0\}} (M_n - M_n \circ \varphi) && (n \in \mathbb{N}) , \end{aligned}$$

hence the desired claim.
2. But this implies:

$$\begin{aligned} \mathbb{E}\left(X\, 1_{\{M_n>0\}} \right) &\geq \int_{\{M_n>0\}} (M_n - M_n \circ \varphi)\, d\mathbb{P} \\ &= \int (M_n - M_n \circ \varphi)\, d\mathbb{P} = 0 , \end{aligned}$$

where we finally use that φ is measure preserving. □

Proof of Birkhoff's ergodic theorem 4.1 Wlog assume $\mathbb{E}(X|\mathscr{I}) = 0$; else we consider $\tilde{X} := X - \mathbb{E}(X|\mathscr{I})$, which is possible thanks to the invariance of $\mathbb{E}(X|\mathscr{I}) \circ \varphi = \mathbb{E}(X|\mathscr{I})$ ($\mathbb{P}$-a.s.).

1. $\mathbb{P}$-almost sure convergence: For this purpose we will show that with

$$\bar{X} \; := \; \limsup_{n\to\infty} \frac{S_n}{n} \; \equiv \; \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X \circ \varphi^k$$

and with

$$D \; := \; \{ \bar{X} > \varepsilon \} \; \in \; \mathscr{I} \qquad \text{(for arbitrary } \varepsilon > 0)$$

we have

$$\mathbb{P}(D) \; = \; 0 \, ;$$

analogously one can show $\liminf \frac{S_n}{n} \geq 0$, by considering $-X$ instead of X. To show $\mathbb{P}(D) = 0$ we give D an alternative description: With

$$\begin{aligned} X^* \; &:= \; (X - \varepsilon)\, 1_D \\ S_n^* \; &:= \; X^* + X^* \circ \varphi + \cdots + X^* \circ \varphi^{n-1} \\ M_n^* \; &:= \; \max\{0, S_1^*, \ldots, S_n^*\} \\ F_n \; &:= \; \{ M_n^* > 0 \} \end{aligned}$$

we have

$$D \; = \; \left\{ \sup_{n\in\mathbb{N}} \frac{S_n^*}{n} \; > \; 0 \right\} \; = \; \bigcup_{n\in\mathbb{N}} F_n \, .$$

Upon applying the maximal-ergodic Lemma 4.2 on X^*, we obtain:

$$\begin{aligned} 0 \quad &\leq \quad \mathbb{E}\,(\, X^* \, 1_{F_n}\,) && \text{(Lemma 4.2)} \\ &\xrightarrow{n\to\infty} \quad \mathbb{E}\left(X^* \, 1_{\bigcup_n F_n} \right) && \text{(dom. conv., since } X \in L^1) \\ &= \quad \mathbb{E}\,(\, X^* \, 1_D\,) && \text{(above char. of } D) \\ &\equiv \quad \mathbb{E}\,(\, X \, 1_D\,) - \varepsilon\, \mathbb{P}(D) && \text{(definition of } X^*) \\ &= \quad -\varepsilon\, \mathbb{P}(D) && (\, \mathbb{E}(X|\mathscr{I}) = 0 \text{ and } D \in \mathscr{I}) \\ &\leq \quad 0\,, \end{aligned}$$

in summary: $\mathbb{P}(D) = 0$.

2. L^1-convergence: For this purpose we truncate X; with some fixed $K > 0$ let

$$X' \; := \; X \, 1_{\{|X| \leq K\}} \qquad \text{and} \qquad X'' \; := \; X - X' \, .$$

The $\mathbb{P}$-a.s. convergence proved above applies in particular to X'; since this convergence is dominated by K, we obtain for X':

$$\frac{1}{n} \sum_{k=0}^{n-1} X' \circ \varphi^k \; \xrightarrow{n\to\infty} \; \mathbb{E}(X'|\mathscr{I}) \quad \text{in } L^1(\mathbb{P})\, .$$

In addition we have

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} X''\circ\varphi^k\right|\right) \;\le\; \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\left(|X''|\circ\varphi^k\right) \;=\; \mathbb{E}(|X''|)\,,$$

where we use that φ is measure preserving; moreover we have by Jensen ($|\,.\,|$ is convex):

$$\mathbb{E}\left(\,\left|\,\mathbb{E}(X''|\mathscr{I})\,\right|\,\right) \;\le\; \mathbb{E}\left(\,\mathbb{E}(\,|X''|\,|\mathscr{I})\,\right) \;=\; \mathbb{E}(|X''|)\,;$$

combining the preceding two inequalities we obtain

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} X''\circ\varphi^k \;-\; \mathbb{E}(X''|\mathscr{I})\right|\right) \;\le\; 2\,\mathbb{E}(|X''|)\,.$$

Let now an arbitrary $\varepsilon > 0$ be fixed; then we may choose $K > 0$ such that

$$2\,\mathbb{E}(|X''|) \;<\; \frac{\varepsilon}{2}$$

(dominated convergence, definition of X''). With the parameters ε and K by the above L^1-convergence for X' we can choose $n_0 \in \mathbb{N}$, such that we have:

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} X'\circ\varphi^k \;-\; \mathbb{E}(X'|\mathscr{I})\right|\right) \;<\; \frac{\varepsilon}{2} \qquad (n \ge n_0)\,.$$

Since now $X \equiv X' + X''$, the preceding three estimates yield:

$$\begin{aligned}
&\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} X\circ\varphi^k \;-\; \mathbb{E}(X|\mathscr{I})\right|\right)\\
&\le\; \mathbb{E}\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} X'\circ\varphi^k - \mathbb{E}(X'|\mathscr{I})\right|\right) + \mathbb{E}\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} X''\circ\varphi^k - \mathbb{E}(X''|\mathscr{I})\right|\right) \;<\; \varepsilon.
\end{aligned}$$

□

Example 4.3 (Strong law of large numbers)**.** Let $(X_n)_{n\in\mathbb{N}_0}$ be i.i.d. random variables, wlog defined on the sequence space $\Omega := \mathbb{R}^{\mathbb{N}_0}$, with $\mathbb{P} \equiv P_X = P_{X_0}\otimes P_{X_0}\otimes\cdots$ and ergodic shift $\varphi = \theta_1$ (see example 3.12). If $X_0 \in L^1(\mathbb{P})$, we infer from Theorem 4.1 with Proposition 3.9:

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \;=\; \frac{1}{n}\sum_{k=0}^{n-1} X_0\circ\varphi^k \;\xrightarrow{\mathbb{P}\text{-a.s.},\,L^1(\mathbb{P})}\; \mathbb{E}(X_0|\mathscr{I}) \;=\; \mathbb{E}(X_0)\,.$$

Example 4.4 (Rotation on the circle, Weyl's equidistribution law)**.** Let

$$\varphi\,:\,\Omega\longrightarrow\Omega\,, \qquad \varphi(\omega) \;:=\; \omega+\theta \;\; (\text{mod } 1)\,,$$

on $(\Omega,\mathscr{F},\mathbb{P}) := ([0,1),\mathscr{B}[0,1),\lambda\big|_{\mathscr{F}})$ be given as in 3.4 and 3.13, where λ is the Lebesgue measure. Moreover let $\theta\in\mathbb{Q}^c$. Then from Theorem 4.1 with example 3.13 for $A\in\mathscr{B}[0,1)$

we obtain:

$$\frac{1}{n}\sum_{k=0}^{n-1} 1_A \circ \varphi^k \xrightarrow{\lambda\text{-a.e.}, L^1(\lambda)} \lambda(A)\,.$$

5. The subadditive ergodic theorem of Kingman

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with a measure preserving transformation $\varphi : \Omega \to \Omega$. In the preceding section we have investigated the asymptotic behavior of $\frac{S_n}{n}$, where S_n is of the form $\sum_{k=0}^{n-1} X \circ \varphi^k$, so in particular satisfies the *additive cocycle property*

$$S_{n+m} \;=\; S_n \;+\; S_m \circ \varphi^n \qquad\qquad (n, m \in \mathbb{N}_0).$$

Now we are interested in the following generalization:

Definition 5.1 (subadditive sequences of random variables)**.** A sequence $(Y_n)_n$ of random variables ($n \in \mathbb{N}_0$ or $\mathbb{N}$; state space $\mathbb{R} \cup \{-\infty\}$) is called *subadditive*, if we have:

$$Y_{n+m} \;\leq\; Y_n \;+\; Y_m \circ \varphi^n \qquad\qquad (n, m \in \mathbb{N}_0)\,.$$

A sequence $(Y_n)_{n\in\mathbb{N}_0}$ is called *superadditive*, if $(-Y_n)_{n\in\mathbb{N}_0}$ is subadditive, and it is called *additive*, if it is both sub- and superadditive.

Example 5.2. Let $(X_n)_{n\in\mathbb{N}_0}$ be a sequence of i.i.d. random variables, realized wlog by $X_n = \pi_{\{n\}}$ on the sequence space $(\Omega, \mathscr{F}, \mathbb{P}, \varphi) = (S^{\mathbb{N}_0}, \mathscr{S}^{\mathbb{N}_0}, P_{(X_n)_{n\in\mathbb{N}_0}}, \theta_1)$. Let

$$S_n \;:=\; \sum_{k=0}^{n-1} X_k\,.$$

Then $(S_n)_{n\in\mathbb{N}_0}$ is additive and $(|S_n|)_{n\in\mathbb{N}_0}$ subadditive.

Proof. Additivity of $(S_n)_n$ is immediate, since $\varphi \equiv \theta_1$. Moreover we have:

$$\begin{aligned} |S_{n+m}| \;\equiv\; \left|\sum_{k=0}^{n+m-1} X_k\right| \;&\leq\; \left|\sum_{k=0}^{n-1} X_k\right| + \left|\sum_{k=n}^{n+m-1} X_k\right| \\ &=\; |S_n| + \left|\sum_{k=0}^{m-1} X_k \circ \varphi^n\right| \;=\; |S_n| + |S_m| \circ \varphi^n\,. \end{aligned}$$

□

Example 5.3. (discrete version of linear stochastic differential equation) Let B_0, B_1 be $d \times d$ matrices with real values, W a one-dimensional Brownian motion, θ_1 the shift by time one on Wiener space. We consider the discrete version

$$x_{n+1} - x_n = B_0 x_n + B_1(W_{n+1} - W_n)$$

of the stochastic differential equation

$$dx_t = B_0 x_t dt + B_1 x_t dW_t.$$

The discrete equation may be written

$$x_{n+1} = (I + B_0 + B_1(W_1 - W_0) \circ \theta^n) x_n = (A \circ \theta^n)\; x_n$$

with the random matrix $A : \Omega \to \mathbb{R}^{d\times d}$ given by $A = (I + B_0 + B_1(W_1 - W_0))$. This is a special case of the following example. On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with measure preserving map φ let a random matrix be given, i.e. a measurable mapping $A : \Omega \to \mathbb{R}^{d\times d}$. Moreover let

$$\begin{aligned} A_n &:= (A\circ\varphi^{n-1})(A\circ\varphi^{n-2})\cdots A && \text{and thus} \\ Y_n &:= \log \| A_n \| && (n\in\mathbb{N})\,, \end{aligned}$$

where $\| \,.\, \|$ denotes a matrix norm. Then $(Y_n)_n$ is subadditive.

Proof.

$$\begin{aligned} Y_{n+m} &= \log \| (A\circ\varphi^{m-1}\circ\varphi^n)\cdots(A\circ\varphi^0\circ\varphi^n)(A\circ\varphi^{n-1})\cdots A \| \\ &= \log \| (A_m\circ\varphi^n)\, A_n \| \\ &\overset{\text{Norm}}{\leq} \log\left[(\| A_m \|\circ\varphi^n)\ \| A_n \| \right] \\ &= \log\,(\| A_m \|\circ\varphi^n) + \log \| A_n \| \quad \equiv \quad Y_m\circ\varphi^n + Y_n\,. \end{aligned}$$

□

Now we aim at getting a convergence statement for subadditive $(Y_n)_n$. This will be achieved in the subadditive ergodic theorem 5.7 of Kingman. For this we state the three following Lemmas.

Lemma 5.4 (Riesz)**.** *Let $u_1, \ldots, u_n \in \mathbb{R}$ $(n\in\mathbb{N})$. With*

$$s_j \ := \ \begin{cases} 0\,, & j=0 \\ u_1+\cdots+u_j\,, & j\in\{1,\ldots,n\}\,, \end{cases}$$

define

$$v_j \ \equiv \ v_{jn} \ := \ \max_{k\in\{j,\ldots,n\}} (\, s_k - s_j \,) \ \equiv \ \max\big\{\, 0\,,\, u_{j+1}\,,\, u_{j+1}+u_{j+2}\,,\, u_{j+1}+\cdots+u_n \,\big\}$$

for $j = 0,1,\ldots,n$. Then we have:

$$\sum_{j=0}^{n-1} u_{j+1}\, 1_{\{v_{jn}>0\}} \ \geq \ 0\,.$$

Proof. 1) First we have for all $j \in \{0, 1, \ldots, n\}$:

$$v_j \;=\; \max\{\, 0 \,,\, u_{j+1} + v_{j+1} \,\} \;\equiv\; (u_{j+1} + v_{j+1})^+ \,.$$

This follows directly, since

$$\begin{aligned} v_j &= \max\{\, 0 \,,\, u_{j+1} \,,\, u_{j+1} + u_{j+2} \,,\, u_{j+1} + \cdots + u_n \,\} \qquad \text{and} \\ &= \max\{0, \max\{u_{j+1}, u_{j+1} + u_{j+2}, \cdots, u_{j+1} + \cdots + u_n\}\,\} \\ &= \max\{0, u_{j+1} + \max\{0, u_{j+2}, u_{j+2} + u_{j+3}, \cdots, u_{j+2} + \cdots + u_n\}\} \\ &= \max\{0, u_{j+1} + v_{j+1}\}. \end{aligned}$$

2) By 1) we have:

$$v_j \;\leq\; v_{j+1} + u_{j+1} 1_{\{v_j > 0\}} \qquad (j \in \{0, 1, \ldots, n\}) \,.$$

Indeed, if $v_j = 0$, this is trivial, and in case $v_j > 0$ we have:

$$0 \;<\; v_j \;\overset{1)}{=}\; (u_{j+1} + v_{j+1})^+ \;\overset{v_j > 0}{=}\; v_{j+1} + u_{j+1} \,.$$

3) From 2) now follows the claim of the Lemma, since:

$$0 \;\leq\; v_0 \;=\; v_0 - v_n \;=\; \sum_{j=0}^{n-1} (v_j - v_{j+1}) \;\overset{2)}{\leq}\; \sum_{j=0}^{n-1} u_{j+1}\, 1_{\{v_j > 0\}} \,.$$

□

In the proof of the subadditive ergodic theorem of Kingman we will compare subadditive sequences $(Y_n)_n$ with additive sequences $X_n = \sum_{i=0}^{n-1} X_0 \circ \varphi^i$. For this purpose we need the following auxiliary argument, for which the preceding Lemma of Riesz will be useful.

Lemma 5.5 (maximal inequality). *Let $(Y_n)_{n\in\mathbb{N}_0}$ be superadditive on $(\Omega,\mathscr{F},\mathbb{P},\varphi)$ and assume $Y_n \geq 0$ for all n. Moreover let $X \geq 0$ be an integrable random variable; set*

$$V \;:=\; \sup_{n\in\mathbb{N}_0}(\,Y_n - X_n\,) - Y_0\,, \qquad \textit{where} \qquad X_n \;:=\; \sum_{i=0}^{n-1} X\circ\varphi^i\,.$$

Then we have:

$$\mathbb{E}\left(X\,1_{\{V>0\}}\,|\,\mathscr{I}\right) \;\leq\; \sup_{n\in\mathbb{N}} \frac{\mathbb{E}(\,Y_n\,|\,\mathscr{I}\,)}{n}\,.$$

Proof. Let $v_{jn} := \max_{k\in\{j,\dots,n\}}\Big(Y_k - Y_j - \sum_{i=j}^{k-1} X\circ\varphi^i\Big)$ for $j = 0,1,\dots,n$.

1) At first we have:

$$Y_n \;\geq\; \sum_{j=0}^{n-1} X\circ\varphi^j\;\; 1_{\{v_{jn}>0\}} \qquad (n\in\mathbb{N})\,;$$

since with $Y_{j+1}\geq Y_j$ (by superadditivity and $Y_n \geq 0$) we obtain:

$$\begin{aligned} Y_n \;\geq\; Y_n - Y_0 \;&=\; \sum_{j=0}^{n-1}(\,Y_{j+1} - Y_j\,) \\ &\geq\; \sum_{j=0}^{n-1}(\,Y_{j+1} - Y_j\,)\;1_{\{v_{jn}>0\}} \;\overset{5.4}{\geq}\; \sum_{j=0}^{n-1} X\circ\varphi^j\;\;1_{\{v_{jn}>0\}}\,, \end{aligned}$$

where the last step follows from Lemma 5.4 with $u_j := Y_j - Y_{j-1} - X\circ\varphi^{j-1}$.

2) From 1) we get:

$$\mathbb{E}(\,Y_n\,|\,\mathscr{I}\,) \;\geq\; \sum_{k=1}^{n} \mathbb{E}\left(X\,1_{\{v_{0k}>0\}}\,|\,\mathscr{I}\right) \qquad (n\in\mathbb{N});$$

since for $k \geq j$ by superadditivity $Y_k - Y_j \;\geq\; Y_{k-j}\circ\varphi^j$ and thus $v_{jn} \geq v_{0(n-j)}\circ\varphi^j$, hence in summary (with the measure preserving property of φ):

$$\begin{aligned} \mathbb{E}(\,Y_n\,|\,\mathscr{I}\,) \;&\overset{1)}{\geq}\; \sum_{j=0}^{n-1} \mathbb{E}\Big(X\circ\varphi^j\;\;1_{\{v_{jn}>0\}}\,\Big|\,\mathscr{I}\Big) \\ &\geq\; \sum_{j=0}^{n-1} \mathbb{E}\Big(\Big[X\;\;1_{\{v_{0(n-j)}>0\}}\Big]\circ\varphi^j\,\Big|\,\mathscr{I}\Big) \;=\; \sum_{k=1}^{n} \mathbb{E}\left(X\,1_{\{v_{0k}>0\}}\,|\,\mathscr{I}\right). \end{aligned}$$

3) By Fatou and $\{v_{0k}>0\} \nearrow \{V>0\}$ $(k\to\infty)$ we obtain from this:

$$\sup_{n\in\mathbb{N}} \frac{\mathbb{E}(\,Y_n\,|\,\mathscr{I}\,)}{n} \;\overset{2)}{\geq}\; \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left(X\,1_{\{v_{0k}>0\}}\,|\,\mathscr{I}\right) \;\geq\; \mathbb{E}\left(X\,1_{\{V>0\}}\,|\,\mathscr{I}\right)\,.$$

□

Lemma 5.6. *If Z is a measurable function on $(\Omega, \mathscr{F}, \mathbb{P}, \varphi)$ with $Z \geq Z \circ \varphi$, we have $Z = Z \circ \varphi$. If in particular $(Y_n)_{n \in \mathbb{N}_0}$ is superadditive and*

$$\overline{Y} := \limsup_{n\to\infty} \frac{Y_n}{n}, \qquad \textit{resp.} \qquad \underline{Y} := \liminf_{n\to\infty} \frac{Y_n}{n},$$

we have:

$$\overline{Y} = \overline{Y} \circ \varphi, \qquad \textit{resp.} \qquad \underline{Y} = \underline{Y} \circ \varphi.$$

PROOF. We first prove the statement on Z and for this assume $Z > Z \circ \varphi$ on a set of positive measure, so

$$\mathbb{P}(Z > q > Z \circ \varphi) > 0$$

for some $q \in \mathbb{Q}$. Then we get:

$$\begin{aligned}
\mathbb{P}(Z < q) &\overset{\varphi \text{ m.p.}}{=} \mathbb{P}(Z \circ \varphi < q) \\
&= \mathbb{P}(Z \circ \varphi < q \leq Z) + \mathbb{P}(Z \circ \varphi < q \,,\, Z < q) \\
&\overset{Z \geq Z\circ\varphi}{=} \underbrace{\mathbb{P}(Z \circ \varphi < q \leq Z)}_{>0} + \mathbb{P}(Z < q) \\
&> \mathbb{P}(Z < q) \,,
\end{aligned}$$

a contradiction.
With what has just been proved, it now remains to show:

$$\overline{Y} \geq \overline{Y} \circ \varphi, \qquad \text{resp.} \qquad \underline{Y} \geq \underline{Y} \circ \varphi;$$

but by the superadditivity of $(Y_n)_n$ we now have:

$$\begin{aligned}
\frac{Y_{n+1}}{n+1} &\geq \frac{Y_1}{n+1} + \frac{Y_n \circ \varphi}{n+1} \\
&= \frac{Y_1}{n+1} + \frac{n}{n+1} \frac{Y_n}{n} \circ \varphi \,.
\end{aligned}$$

□

Theorem 5.7 (subadditive ergodic theorem, Kingman).
On $(\Omega, \mathscr{F}, \mathbb{P}, \varphi)$ let $(Y_n)_{n \in \mathbb{N}}$ be a superadditive sequence of integrable random variables. Then we have:

$$\frac{Y_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}\text{-}a.s.} \sup_{n\in\mathbb{N}} \frac{1}{n}\, \mathbb{E}(Y_n \,|\, \mathscr{I}) =: \gamma \leq \infty \,.$$

Here γ is integrable iff $\sup_{n\in\mathbb{N}} \frac{1}{n}\, \mathbb{E}(Y_n) < \infty$. In this case we also have

$$\frac{Y_n}{n} \xrightarrow[n\to\infty]{L^1(\mathbb{P})} \gamma \,.$$

Moreover there exists a set $\tilde{\Omega} \in \mathscr{I}$ with $\tilde{\Omega} \subset \varphi^{-1}\, \tilde{\Omega}$ and $\mathbb{P}(\tilde{\Omega}) = 1$, so that we also have:

$$\frac{Y_n}{n} \xrightarrow[n\to\infty]{} \gamma \quad \textit{on} \quad \tilde{\Omega} \,.$$

Proof. To simplify notation we set $Y_0 := 0$ (else replace Y_n by $Y_n - Y_0$, $n \geq 0$); then $(Y_n)_{n\in\mathbb{N}_0}$ is furthermore superadditive.

1) We first show that we can assume wlog

$$Y_n \geq 0 \qquad (n \in \mathbb{N}).$$

For this purpose let

$$G_n := Y_n - F_n, \qquad \text{with} \qquad F_n := \sum_{i=0}^{n-1} Y_1 \circ \varphi^i .$$

Then

$$Y_n = G_n + F_n.$$

The sequence $(F_n)_n$ is an additive cocycle, so that the assumed integrability provides due to Birkhoffs ergodic theorem 4.1 a random variable γ_F such that

$$\frac{F_n}{n} \longrightarrow \gamma_F$$

$\mathbb{P}$-a.s. and in $L^1(\mathbb{P})$. The claims about Y follow, if we have shown

$$\frac{G_n}{n} \longrightarrow \gamma_G \qquad (\mathbb{P}\text{-a.s. and in } L^1(\mathbb{P})),$$

since then in particular

$$\frac{Y_n}{n} \longrightarrow \gamma_G + \gamma_F \qquad (\mathbb{P}\text{-a.s. and in } L^1(\mathbb{P})).$$

But now by inductive application of superadditivity

$$\begin{aligned} G_n &= -Y_1 - Y_1 \circ \varphi - \cdots - Y_1 \circ \varphi^{n-2} - Y_1 \circ \varphi^{n-1} + Y_n \\ &\geq -Y_1 - Y_1 \circ \varphi - \cdots - Y_1 \circ \varphi^{n-2} + Y_{n-1} \\ &\geq -Y_1 - Y_1 \circ \varphi - \cdots + Y_{n-2} \\ &\geq \cdots\cdots \\ &\geq -Y_1 - Y_1 \circ \varphi + Y_2 \geq 0\,; \end{aligned}$$

on the other hand superadditivity of $(Y_n)_n$ is transferred to $(G_n)_n$. Therefore the claims for $(Y_n)_n$ reduce to corresponding convergences for the positive process $(G_n)_n$.

2) We further prove that wlog we can assume

$$Y_n \geq n \qquad (n \in \mathbb{N}) :$$

Since $Y_{n+m} + n + m \geq Y_n + n + (Y_m + m) \circ \varphi^n$ with $(Y_n)_n$ also $(Y_n + n)_n$ is superadditive. If $\frac{Y_n+n}{n} \longrightarrow \gamma'$, so also $\frac{Y_n}{n} \longrightarrow \gamma := \gamma' - 1$.

3) $\frac{Y_n}{n} \longrightarrow \gamma$ $\mathbb{P}$-a.s.: For this we show $\overline{Y} \le \gamma$ and $\underline{Y} \ge \gamma$, where again

$$\overline{Y} := \limsup_{n\to\infty} \frac{Y_n}{n} \qquad \text{resp.} \qquad \underline{Y} := \liminf_{n\to\infty} \frac{Y_n}{n}.$$

i) $\overline{Y} \le \gamma$ $\mathbb{P}$-a.s.: For $2 < r \in \mathbb{N}$ we define

$$X^r := \min\{ r, \overline{Y} - \frac{1}{r} \} > 0;$$

with this the inequality follows, since by 2) $\overline{Y} \ge 1$. By Lemma 5.6 we have moreover $X^r = X^r \circ \varphi$, hence $X^r_n := \sum_{i=0}^{n-1} X^r \circ \varphi^i = nX^r$; with this we have

$$V := \sup_{n\in\mathbb{N}_0} (Y_n - X^r_n) - Y_0 = \sup_{n\in\mathbb{N}_0} (Y_n - nX^r) > 0;$$

here the latter inequality follows from the definition of X^r: in case $Y_n \le nX^r$ for all $n \in \mathbb{N}_0$, the contradiction $\overline{Y} \equiv \limsup \frac{Y_n}{n} \le \limsup \frac{nX^r}{n} = X^r < \overline{Y}$ would follow. Hence with Lemma 5.5:

$$X^r = \mathbb{E}(X^r \,|\, \mathscr{I}) \le \sup_{n\in\mathbb{N}} \frac{\mathbb{E}(Y_n \,|\, \mathscr{I})}{n} \equiv \gamma.$$

But from this as $r \to \infty$: $\overline{Y} \le \gamma$ $\mathbb{P}$-a.s..

ii) $\underline{Y} \ge \gamma$ $\mathbb{P}$-a.s.: First we have that $(Y_n)_n$ is increasing, by superadditivity and positivity; from this we conclude:

$$k\, Y_{n+k-1} \ge \sum_{l=1}^{k} \sum_{j=0}^{n-1} (Y_{j+l} - Y_{j+l-1}) = \sum_{j=0}^{n-1} (Y_{j+k} - Y_j) \qquad (k, n \in \mathbb{N});$$

consequently for each $k \in \mathbb{N}$:

$$\begin{aligned}
\underline{Y} &= \liminf_{n\to\infty} \frac{Y_{n+k-1}}{n+k-1} \\
&= \liminf_{n\to\infty} \frac{Y_{n+k-1}}{n} \\
&= \frac{1}{k} \liminf_{n\to\infty} \frac{k\, Y_{n+k-1}}{n} \\
&\ge \frac{1}{k} \liminf_{n\to\infty} \sum_{j=0}^{n-1} \frac{Y_{j+k} - Y_j}{n} && \text{(prec. rem.)} \\
&\ge \frac{1}{k} \liminf_{n\to\infty} \sum_{j=0}^{n-1} \frac{Y_k \circ \varphi^j}{n} && \text{(superadditivity)} \\
&= \frac{1}{k} \mathbb{E}(Y_k \,|\, \mathscr{J}) && \text{(Birkhoff 4.1)}
\end{aligned}$$

hence also $\underline{Y} \ge \gamma$.

4) γ integrable $\Leftrightarrow \sup\limits_{n\in\mathbb{N}} \frac{1}{n}\,\mathbb{E}(Y_n) < \infty$: For this let $Z_n := \frac{Y_n}{n}$. From what has been shown: $Z_n \to \gamma$ $\mathbb{P}$-a.s. and $\mathbb{E}\gamma \geq \mathbb{E}Z_n$; therefore "'$\Rightarrow$"' has been shown. "'$\Leftarrow$"' follows with monotone convergence.

5) If γ is integrable, we have: $Z_n \equiv \frac{Y_n}{n} \to \gamma$ in $L^1(\mathbb{P})$.
In fact, by $0 \leq (\gamma - Z_n)^+ \leq \gamma$ we have on the one hand

$$\mathbb{E}\left((\gamma - Z_n)^+ \right) \to 0\,;$$

and on the other hand also

$$0 \leq \mathbb{E}(\gamma - Z_n) \to 0$$

by Fatou, so that we also obtain:

$$\mathbb{E}\left((\gamma - Z_n)^- \right) = -\mathbb{E}(\gamma - Z_n) + \mathbb{E}\left((\gamma - Z_n)^+ \right) \to 0\,,$$

and thus in summary

$$\mathbb{E}\left(|\gamma - Z_n| \right) \to 0\,.$$

6) existence of $\tilde{\Omega} \in \mathscr{I}$ with $\tilde{\Omega} \subset \varphi^{-1}\tilde{\Omega}$, $\mathbb{P}(\tilde{\Omega}) = 1$ and $\frac{Y_n}{n} \to \gamma$ on $\tilde{\Omega}$: By Lemma 5.6 $\overline{Y}$ and $\underline{Y}$ are invariant. Therefore also

$$\tilde{\Omega} := \{ \overline{Y} = \underline{Y} \}$$

is invariant; the remaining properties follow from what has been proved (evtl. choice of an a.s. equal strictly invariant set). □

6. The theorem of Furstenberg-Kesten

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with a measure preserving mapping $\varphi : \Omega \to \Omega$ and $A : \Omega \to \mathbb{R}^{d\times d}$ a random matrix. We now investigate the asymptotics of

$$A_n(\omega) \;:=\; \big(A\circ\varphi^{n-1}(\omega)\big)\,\big(A\circ\varphi^{n-2}(\omega)\big)\cdots\big(A\circ\varphi(\omega)\big)\,\big(A(\omega)\big) \qquad (\omega\in\Omega)\,. \qquad (2)$$

Example 6.1 (deterministic, symmetric matrix)**.** Let $A \in \mathbb{R}^{d\times d}$ be symmetric. Then there exists a diagonalisation of A with real (A symmetric) eigenvalues $\delta_1 \geq \cdots \geq \delta_d$; hence there exists an orthogonal matrix O, so that we have

$$A \;=\; O^*\,DO \qquad \text{with} \qquad D := \begin{pmatrix} \delta_1 & & 0 \\ & \ddots & \\ 0 & & \delta_d \end{pmatrix}.$$

Here we assume $\delta_1 > \cdots > \delta_d$, i.e. the eigenspaces E_i belonging to δ_i are one dimensional. Moreover let x_i be an eigenvector in E_i and

$$V_j \;=\; \begin{cases} E_j \oplus E_{j+1} \oplus \cdots \oplus E_d\,, & j = 1,\ldots,d\,, \\ \{0\}\,, & j = d+1\,. \end{cases}$$

In this setting let $x \in V_j \setminus V_{j+1}$. Then x can be written as

$$x \;=\; \sum_{k=j}^{d} \alpha_k\, x_k \qquad \text{with} \qquad \alpha_j \neq 0\,.$$

So we have by linearity of A, since x_k are eigenvectors:

$$A^n x \;=\; \sum_{k=j}^{d} \alpha_k\, A^n x_k \;=\; \sum_{k=j}^{d} \alpha_k\, \delta_k^n\, x_k\,,$$

hence

$$\begin{aligned} \frac{1}{n}\log|A^n x| \;&=\; \frac{1}{n}\log\left|\sum_{k=j}^{d}\alpha_k\,\delta_k^n\,x_k\right| \\ &=\; \frac{1}{n}\left[\log\delta_j^n \,+\, \log\left|\sum_{k=j}^{d}\alpha_k\left(\frac{\delta_k}{\delta_j}\right)^n x_k\right|\right] \;\xrightarrow{n\to\infty}\; \log\delta_j\,. \end{aligned}$$

We also have the reverse conclusion, in summary:

$$x \in V_j\setminus V_{j+1} \quad\Longleftrightarrow\quad \lim_{n\to\infty}\frac{1}{n}\log\|A^n x\| \;=\; \log\delta_j \qquad (j=1,\ldots,d)\,.$$

We now aim at showing this analogously also for the sequence $(A^n)_n$ from (2).

Definition-remark 6.2 (decomposition according to singular value)**.** Each $A \in \mathbb{R}^{d\times d}$ possesses a *singular value decomposition*, i.e. there are orthogonal matrices U, V and a diagonal matrix

$$D = \begin{pmatrix} \delta_1 & & 0 \\ & \ddots & \\ 0 & & \delta_d \end{pmatrix}$$

with $\delta_1 \geq \cdots \geq \delta_d$, such that we have

$$A = VDU\,.$$

Thereby $\delta_1, \ldots, \delta_d$ are the eigenvalues of $(A^*A)^{1/2}$ and for the operator norm we have: $\|A\| = \delta_1$.

Proof. First of all, A possesses a *polar decomposition*, i.e.:

$$A = W(A^*A)^{1/2} \quad \text{with an orthogonal matrix } W\,.$$

(In case A is non-singular, this follows with $W := A(A^*A)^{-1/2}$). Let now $D := \operatorname{diag}(\delta_1, \ldots, \delta_d)$ be the diagonal matrix with eigenvalues $\delta_1 \geq \cdots \geq \delta_d$ of $(A^*A)^{1/2}$. So the positive semidefinite matrix $(A^*A)^{1/2}$ can be written as

$$(A^*A)^{1/2} = U^*DU$$

with an orthogonal matrix U. In summary we have with $V := WU^*$

$$A = W(A^*A)^{1/2} = WU^*DU = VDU\,.$$

□

Remark: If $e_1, \ldots, e_d$ are the canonical unit vectors in $\mathbb{R}^d$, then Ue_i is the vector in direction of the i-th main axis of the ellipsoid $(A^*A)^{1/2}(S^{d-1})$ and δ_i describes the dilation in this direction.

For the construction of the analogues of $\delta_1, \ldots, \delta_d$ in example 6.1 for the sequence A_n defined in (2) we need information on how A_n acts on linear subspaces of $\mathbb{R}^d$ of any dimension below d. These linear subspaces are elements of the Grassmannian manifolds which can be defined via exterior products. In the Theorem by Furstenberg-Kesten we shall investigate the asymptotic behavior of the singular values $\delta_i(A_n)$ for $1 \leq i \leq d$ as $n \to \infty$.

Definition 6.3 (exterior product)**.** For a d-dimensional linear space E let E^* be the dual of E and

$$\mathscr{L}^k(E^*) := \{\,k\text{-linear forms on } (E^*)^k\,\} \qquad (k = 1, \ldots, d)\,.$$

We thus define $\wedge^k E$, the *k-fold "exterior product of* E", as

$$\wedge^k E := \{\, f \in \mathscr{L}^k(E^*) \,:\, f \text{ alternating}\,\}\,,$$

hence as the collection of all k-linear, alternating multilinear forms on $(E^*)^k$.
An element $f \in \wedge^k E$ is a k-linear mapping

$$f : \underbrace{E^* \times \cdots \times E^*}_{k\text{times}} \longrightarrow \mathbb{R}\,,$$

which is alternating, i.e.:

$$f(\,\dots\,, x_i\,,\,\dots\,, x_j\,,\,\dots\,) \;=\; -\,f(\,\dots\,, x_j\,,\,\dots\,, x_i\,,\,\dots\,) \qquad (i \neq j)\,.$$

Lemma 6.4 (alternating maps)**.** *For $f \in \mathscr{L}^k(E^*)$ the following are equivalent:*

i) $f \in \wedge^k E$;

ii) $f(x_1\,,\,\dots\,,x_k) \;=\; 0$, *if* $(x_1\,,\,\dots\,,x_k)$ *not pairwise different;*

iii) $f(x_1\,,\,\dots\,,x_k) \;=\; 0$, *if* $(x_1\,,\,\dots\,,x_k)$ *not pairwise linearly independent;*

iv) $f(x_{\pi(1)}\,,\,\dots\,,x_{\pi(k)}) \;=\; sgn(\pi)\, f(x_1\,,\,\dots\,,x_k)$ *for all* $\pi \in \mathfrak{S}_k$.

Proof. i)⇔iv) follows from the representation $\pi \;=\; \tau_1 \circ \cdots \circ \tau_k$ with permutations of two elements τ_i , hence $\operatorname{sgn}(\tau_i) = -1$.

i)⇒ii): If $(x_1\,,\,\dots\,,x_k)$ are not pairwise different, from the definition of an alternating mapping we obtain by exchanging equal elements: $f(x_1\,,\,\dots\,,x_k) \;=\; -f(x_1\,,\,\dots\,,x_k)$.

ii)⇒iii): Let wlog $x_k \;=\; \sum_{i=1}^{k-1} \alpha_i\, x_i$. Then with linearity and ii): $f(x_1\,,\,\dots\,,x_k) \;=\; \sum_{i=1}^{k-1}\; \alpha_i\, f(x_1\,,\,\dots\,,x_{k-1}\,,x_i) \;=\; 0$.

iii)⇒ii) is trivial.

ii)⇒i): Let wlog $k = 2$. Then we have for $x_1, x_2 \in E^*$:

$$\begin{aligned} 0 &\overset{\text{ii)}}{=} f(x_1 + x_2\,,\, x_1 + x_2) \\ &= f(x_1, x_1) \,+\, f(x_1, x_2) \,+\, f(x_2, x_1) \,+\, f(x_2, x_2) \\ &\overset{\text{ii)}}{=} f(x_1, x_2) \,+\, f(x_2, x_1)\,, \end{aligned}$$

hence $f(x_1, x_2) \;=\; -f(x_2, x_1)$. □

Definition-remark 6.5. Let $f \in \wedge^k E$ and $g \in \wedge^l E$, where E denotes again a d-dimensional linear space and $k, l \in \mathbb{N}_0$. Then

$$f \wedge g\,(\,x_1\,,\,\dots\,,x_{k+l}\,) \;:=\; \frac{1}{k!\,l!} \sum_{\pi \in \mathfrak{S}_{k+l}} \operatorname{sgn}(\pi)\, f\left(x_{\pi(1)}\,,\,\dots\,,x_{\pi(k)}\right)\; g\left(x_{\pi(k+1)}\,,\,\dots\,,x_{\pi(k+l)}\right)$$

is called the *exterior product of* f *and* g and we have: $f \wedge g \in \wedge^{k+l} E$.

Proof. We have: $f \wedge g \in \mathscr{L}^m(E^*)$ with $m := k + l$. To see that $f \wedge g$ is alternating, we apply 6.4 iv); for arbitrary $x_1, \ldots, x_m \in E^*$ let

$$a(\pi) := \frac{1}{k!\, l!} f\left(x_{\pi(1)}, \ldots, x_{\pi(k)}\right) g\left(x_{\pi(k+1)}, \ldots, x_{\pi(m)}\right) \qquad (\pi \in \mathfrak{S}_m).$$

Then:

$$\begin{aligned} f \wedge g\,(x_{\pi(1)}, \ldots, x_{\pi(m)}) &= \sum_{\pi' \in \mathfrak{S}_m} \operatorname{sgn}(\pi')\, a(\pi' \circ \pi) \\ &= \operatorname{sgn}(\pi) \sum_{\pi' \in \mathfrak{S}_m} \operatorname{sgn}(\pi' \circ \pi)\, a(\pi' \circ \pi) \\ &= \operatorname{sgn}(\pi) \sum_{\sigma \in \mathfrak{S}_m} \operatorname{sgn}(\sigma)\, a(\sigma) \\ &= \operatorname{sgn}(\pi)\, f \wedge g\,(x_1, \ldots, x_m). \end{aligned}$$

□

Lemma 6.6 (associativity of the exterior product).
Let $f \in \wedge^k E$, $g \in \wedge^l E$ and $h \in \wedge^m E$ with $k, l, m \in \mathbb{N}_0$. Then we have:

$$(f \wedge g) \wedge h = f \wedge (g \wedge h).$$

Proof. Let $n := k + l + m$ and $\mathfrak{T} := \{\tau \in \mathfrak{S}_n : \tau(i) = i \text{ for } i > k + l\}$; moreover let for arbitrary $x_1, \ldots, x_m \in E^*$ and $\pi \in \mathfrak{S}_m$

$$a(\pi) := f\left(x_{\pi(1)}, \ldots, x_{\pi(k)}\right) g\left(x_{\pi(k+1)}, \ldots, x_{\pi(k+l)}\right) h\left(x_{\pi(k+l+1)}, \ldots, x_{\pi(n)}\right).$$

Herewith by twice applying Remark 6.5:

$$\begin{aligned} &\big((f \wedge g) \wedge h\big)(x_1, \ldots, x_n) = \\ &\quad = \frac{1}{(k+l)!\, m!} \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) \frac{1}{k!\, l!} \sum_{\tau \in \mathfrak{T}} \operatorname{sgn}(\tau)\, a(\sigma \circ \tau) \\ &\quad = \frac{1}{k!\, l!\, m!} \frac{1}{(k+l)!} \sum_{\tau \in \mathfrak{T}} \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma \circ \tau)\, a(\sigma \circ \tau) \\ &\quad = \frac{1}{k!\, l!\, m!} \frac{\operatorname{card}(\mathfrak{T})}{(k+l)!} \sum_{\gamma \in \mathfrak{S}_n} \operatorname{sgn}(\gamma)\, a(\gamma) \\ &\quad = \frac{1}{k!\, l!\, m!} \sum_{\gamma \in \mathfrak{S}_n} \operatorname{sgn}(\gamma)\, a(\gamma). \end{aligned}$$

Since we obtain this result also, if we (with identical steps) calculate $(f \wedge (g \wedge h))(x_1, \ldots, x_n)$, the claim is proven. □

Hence it is clear that expressions such as

$$f_1 \wedge \ldots \wedge f_m \qquad \text{with } f_l \in \wedge^{k_l} E$$

are uniquely determined. Thus we have:

Lemma 6.7. *Let $f_l \in \wedge^{k_l} E$ for $l \in \{1, \dots, m\}$. Then we have with $n := k_1 + \cdots + k_m$*

$$f_1 \wedge \ldots \wedge f_m \;=\; \prod_{1 \le l \le m} \frac{1}{k_l\,!} \cdot \sum_{\pi \in \mathfrak{S}_n} sgn(\pi)\, f_\pi \,,$$

where f_π with $i_l \;:=\; k_1 + \cdots + k_{l-1}$ is defined as

$$f_\pi(x_1, \dots, x_n) := f_1\left(x_{\pi(1)}, \dots, x_{\pi(i_1)}\right)\, f_2\left(x_{\pi(i_1+1)}, \dots, x_{\pi(i_2)}\right) \cdots f_m\left(x_{\pi(i_m+1)}, \dots, x_{\pi(n)}\right).$$

Proof. This follows by induction on m. The case $l = 2$ is just the definition of 6.5; the case $l = 3$ is shown in the proof of 6.6. □

Lemma 6.8. *Let $e_1, \dots, e_d$ be a basis of $E^{**} \cong E$ and $b_1, \dots, b_d$ the dual basis of E^*. Then we have for all $f \in \wedge^k E$:*

$$f = \sum_{i_1 < \cdots < i_k} a_{i_1 \dots i_k}\, e_{i_1} \wedge \ldots \wedge e_{i_k} \quad \Longleftrightarrow \quad a_{i_1 \dots i_k} \;=\; f(b_{i_1}, \dots, b_{i_k}) \;\; \textit{for all } i_1 < \cdots < i_k \,.$$

Proof. Note first that for $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_k$ from Lemma 6.7:

$$\begin{aligned} e_{i_1} \wedge \ldots \wedge e_{i_k}\,(b_{j_1}, \dots, b_{j_k}) \;&=\; \sum_{\pi \in \mathfrak{S}_k} \mathrm{sgn}(\pi)\; e_{i_1}\left(b_{j_{\pi(1)}}\right) \cdots\; e_{i_k}\left(b_{j_{\pi(k)}}\right) \\ &=\; \sum_{\pi \in \mathfrak{S}_k} \mathrm{sgn}(\pi)\;\; \delta_{i_1\,, j_{\pi(1)}} \cdots\; \delta_{i_k\,, j_{\pi(k)}} \\ &=\; \delta_{i_1\,, j_1} \cdots\; \delta_{i_k\,, j_k} \qquad\qquad (\text{da } i_1 < \cdots < i_k) \\ &=\; \begin{cases} 1\,, & \text{if } i_1 = j_1\,,\, \dots\,,\, i_k = j_k \\ 0\,, & \text{else.} \end{cases} \end{aligned}$$

"'⇒"' follows directly from this remark;
"'⇐"' Let $g := \sum_{i_1 < \cdots < i_k} f(b_{i_1}, \dots, b_{i_k})\, e_{i_1} \wedge \ldots \wedge e_{i_k} \;\in \wedge^k E$. By above remark we have $f(b_{i_1}, \dots, b_{i_k}) = g(b_{i_1}, \dots, b_{i_k})$ for all $i_1 < \cdots < i_k$; by linearity therefore f and g are equal on $(E^*)^k$. □

Lemma 6.9. *Let $e_1, \dots, e_d$ be a basis of $E^{**} \cong E$ and $k \in \{1, \dots, d\}$. Then*

$$\{\, e_{i_1} \wedge \ldots \wedge e_{i_k} \;:\; 1 \le i_1 < \ldots < i_k \le d \,\}$$

is a basis of $\wedge^k E$. In particular we have:

$$\dim \wedge^k E = \binom{d}{k} \,.$$

Proof. To prove this, we choose a basis $b_1, \dots, b_d$ of E^* dual to $e_1, \dots, e_d$ and apply 6.8. □

Definition 6.10 (scalar product)**.** Let $b_1, \ldots, b_d$ be a basis of E^*. Then

$$\langle f, g\rangle := \sum_{i_1<\cdots<i_k} f(b_{i_1}, \ldots, b_{i_k})\, g(b_{i_1}, \ldots, b_{i_k}) \qquad (f, g \in \wedge^k E)$$

defines a scalar product on $\wedge^k E$.

Lemma 6.11. *For the scalar product from 6.10 we have:*

$$\langle u_1 \wedge \ldots \wedge u_k\, ,\, v_1 \wedge \ldots \wedge v_k\rangle = \det\big(\langle u_i, v_j\rangle\big)_{1\le i,j\le k} \qquad (u_i, v_i \in E)\ .$$

In particular we have for the associated norm:

$$|\, u_1 \wedge \ldots \wedge u_k\,| = \sqrt{\det\big(\langle u_i, u_j\rangle\big)_{1\le i,j\le k}} \qquad (u_i, v_i \in E)\ .$$

Proof. The right hand side $h(u_1, \ldots, u_k; v_1, \ldots, v_k) := \det\big(\langle u_i, v_j\rangle\big)_{1\le i,j\le k}$ for fixed $v_1, \ldots, v_k$ is an alternating multilinear form in $u_1, \ldots, u_k$ and vice versa, i.e. $h(\ .\ ; v_1, \ldots, v_k) \in \wedge^k E^*$ and $h(u_1, \ldots, u_k;\ .\ ) \in \wedge^k E^*$. If $e_1, \ldots, e_d$ denotes the dual basis of $b_1, \ldots, b_d$ of E, and if we apply Lemma 6.8 twice, we obtain:

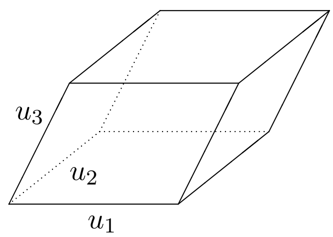
$$\begin{aligned}
&h(u_1, \ldots, u_k; v_1, \ldots, v_k)\\
&\quad= \sum_{i_1<\cdots<i_k} h(e_{i_1}, \ldots, e_{i_k}; v_1, \ldots, v_k)\, b_{i_1} \wedge \ldots \wedge b_{i_k}(u_1, \ldots, u_k)\\
&\quad= \sum_{i_1<\cdots<i_k} \sum_{j_1<\cdots<j_k} h(e_{i_1}, \ldots, e_{i_k}; e_{j_1}, \ldots, e_{j_k})\, b_{i_1} \wedge \ldots \wedge b_{i_k}(u_1, \ldots, u_k)\\
&\qquad\qquad \cdot\, b_{j_1} \wedge \ldots \wedge b_{j_k}(v_1, \ldots, v_k)\\
&\quad= \sum_{i_1<\cdots<i_k} b_{i_1} \wedge \ldots \wedge b_{i_k}(u_1, \ldots, u_k)\ \ b_{i_1} \wedge \ldots \wedge b_{i_k}(v_1, \ldots, v_k)\\
&\quad= \sum_{i_1<\cdots<i_k} u_1 \wedge \ldots \wedge u_k(b_{i_1}, \ldots, b_{i_k})\ \ v_1 \wedge \ldots \wedge v_k(b_{i_1}, \ldots, b_{i_k})\\
&\quad\equiv \langle u_1 \wedge \ldots \wedge u_k\, ,\, v_1 \wedge \ldots \wedge v_k\rangle\ .
\end{aligned}$$

□

Remark: In particular the preceding statement holds for $E := \mathbb{R}^d = E^*$ (with the canonical basis). The norm

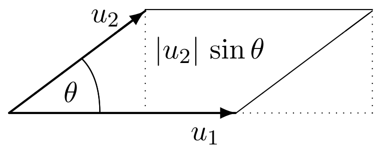
$$|\, u_1 \wedge \ldots \wedge u_k\,| = \sqrt{\det\big(\langle u_i, u_j\rangle\big)_{1\le i,j\le k}}$$

provides for $E = \mathbb{R}^d$ the k-dimensional volume of the paralellipiped spanned by $u_1, \ldots, u_k$.



So we have e.g. in the case $k = 2$:

$$
\begin{aligned}
|\, u_1 \wedge u_2 \,| &= \det \begin{pmatrix} \langle u_1, u_1 \rangle & \langle u_1, u_2 \rangle \\ \langle u_2, u_1 \rangle & \langle u_2, u_2 \rangle \end{pmatrix}^{1/2} \\
&= \det \begin{pmatrix} |u_1|^2 & |u_1|\,|u_2|\, \cos\theta \\ |u_1|\,|u_2|\, \cos\theta & |u_2|^2 \end{pmatrix}^{1/2} \\
&= \left(|u_1|^2\, |u_2|^2\, (1 - \cos^2\theta) \right)^{1/2} \\
&= |u_1|\, |u_2|\, |\sin\theta|
\end{aligned}
$$



□

Since we want to study the action of A on k-dimensional objects in $\mathbb{R}^d$, we now consider the k-fold exterior product of a matrix:

Definition-remark 6.12. Let $A \in \mathbb{R}^{d\times d}$. Then by Lemma 6.9 via

$$\wedge^k A\,(u_1 \wedge \ldots \wedge u_k) \;:=\; Au_1 \wedge \ldots \wedge Au_k \qquad\qquad (u_i \in \mathbb{R}^d)$$

a linear operator $\wedge^k A : \wedge^k\mathbb{R}^d \to \wedge^k\mathbb{R}^d$ is defined, the *k-fold exterior product of the matrix A*. For this we have:

i) $\wedge^1 A = A$,

ii) $\wedge^d A = \det A$ (by Lemma 6.7),

iii) $\wedge^k(AB) = (\wedge^k A)(\wedge^k B)$,

iv) $(\wedge^k A)^{-1} = \wedge^k A^{-1}$ if A invertible,

v) $\wedge^k(cA) = c^k \wedge^k A$ for $c \in \mathbb{R}$,

vi) $\wedge^k U$ orthogonal, if U orthogonal and in this case we have $(\wedge^k U)^* = \wedge^k U^*$.

Lemma 6.13 (exterior product of a matrix and eigenvalues)**.** *Let $\lambda_1, \ldots, \lambda_d$ be the eigenvalues of $A \in \mathbb{R}^{d\times d}$. Then $\wedge^k A$ has the eigenvalues*

$$\{\, \lambda_{i_1} \cdots \lambda_{i_k} \;:\; 1 \le i_1 < \cdots < i_k \le d \,\}\,.$$

Proof. If $u_1, \ldots, u_d$ are the eigenvectors of $\lambda_1, \ldots, \lambda_d$, fixing indices $1 \le i_1 < \cdots < i_k \le d$, we get:

$$\begin{aligned} \wedge^k A\,(u_{i_1} \wedge \ldots \wedge u_{i_k}) \;&\equiv\; Au_{i_1} \wedge \ldots \wedge Au_{i_k} \\ &=\; \lambda_{i_1} u_{i_1} \wedge \ldots \wedge \lambda_{i_k} u_{i_k} \\ &=\; (\lambda_{i_1} \cdots \lambda_{i_k})\,(u_{i_1} \wedge \ldots \wedge u_{i_k})\,, \end{aligned}$$

so that $\lambda_{i_1} \cdots \lambda_{i_k}$ is an eigenvalue with eigenvector $u_{i_1} \wedge \ldots \wedge u_{i_k}$. For dimension reasons these are all eigenvectors and hence all eigenvalues. ☐

Lemma 6.14 (exterior product of a matrix and singular value decomposition)**.**
For $A \in \mathbb{R}^{d\times d}$ let $\delta_1 \ge \ldots \ge \delta_d \ge 0$ be the singular values and

$$A = VDU$$

a singular value decomposition, where $D \equiv \operatorname{diag}(\delta_1, \ldots, \delta_d)$. Then we have for $k = 1, \ldots, d$:

i) $\wedge^k A = (\wedge^k V)(\wedge^k D)(\wedge^k U)$ *ist singular value decomposition of $\wedge^k A$;*

ii) $\wedge^k D = \operatorname{diag}(\, \delta_{i_1} \cdots \delta_{i_k} \;:\; 1 \le i_1 < \cdots < i_k \le d\,)$.
Hence $\delta_1 \cdots \delta_k$ is the biggest resp. $\delta_{d-k+1} \cdots \delta_d$ the smallest singular value of $\wedge^k A$.

iii) For the operator norm we have:
$\| \wedge^k A\| = \delta_1 \cdots \delta_k$, $\;|\det A| = \| \wedge^d A\| = \delta_1 \cdots \delta_d$ *and* $\;\| \wedge^k A\| \le \|A\|^k$.

Proof. i) and ii) follow from Remark 6.12 and Lemma 6.13; iii) follows from ii) and the definition of the operator norm $\|\cdot\|$. ☐

Theorem 6.15 (Furstenberg-Kesten)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $A : \Omega \to \mathbb{R}^{d\times d}$ a random matrix, for which we have*

$$\log^+ \| A(\,.\,) \| \in L^1(\mathbb{P})\,. \tag{3}$$

Moreover let as in (2)

$$A_n \;:=\; \big(A\circ\varphi^{n-1}\big)\,\big(A\circ\varphi^{n-2}\big)\cdots(A\circ\varphi)\;A$$

with a ($\mathbb{P}$-)measure preserving map $\varphi : \Omega \to \Omega$.

Then there exists a set $\tilde\Omega \in \mathscr{F}$ with $\mathbb{P}(\tilde\Omega) = 1$ and $\tilde\Omega \subset \varphi^{-1}(\tilde\Omega)$, and there exist measurable functions

$$\gamma^{(k)} : \Omega \longrightarrow \mathbb{R}\cup\{-\infty\} \qquad\qquad (k = 1,\ldots,d)$$

with $\gamma^{(k)+} \in L^1(\mathbb{P})$, such that for all $\omega \in \tilde\Omega$ and $k, m \in \{1,\ldots,d\}$ we have:

$$\begin{aligned}
\gamma^{(k)}(\omega) &= \lim_{n\to\infty}\frac{1}{n}\log\|\wedge^k A_n(\omega)\|\,,\\
\gamma^{(k)}\big(\varphi(\omega)\big) &= \gamma^{(k)}(\omega)\,,\\
\gamma^{(k+m)}(\omega) &\le \gamma^{(k)}(\omega) + \gamma^{(m)}(\omega)\,.
\end{aligned}$$

If we define recursively random variables

$$\Lambda_k : \Omega \longrightarrow \mathbb{R}\cup\{-\infty\} \qquad\qquad (k = 1,\ldots,d)$$

by

$$\Lambda_1 + \ldots + \Lambda_k \;=\; \gamma^{(k)}$$

with

$$\Lambda_k \;:=\; -\infty \quad on \quad \{\,\gamma^{(k)} = -\infty\,\}\,,$$

we have for all $\omega \in \tilde\Omega$ and $k \in \{1,\ldots,d\}$:

$$\begin{aligned}
\Lambda_k(\omega) &= \lim_{n\to\infty}\frac{1}{n}\log\delta_k\big(A_n(\omega)\big)\,,\\
\Lambda_k\big(\varphi(\omega)\big) &= \Lambda_k(\omega)\,,\\
\Lambda_1(\omega) &\ge \Lambda_2(\omega) \;\ge\; \ldots \;\ge\; \Lambda_d(\omega) \;\;(\ge -\infty\,)\,.
\end{aligned}$$

If $\mathbb{P}$ is ergodic, so $\gamma^{(k)}$ and Λ_k are constant by the invariance above (on $\tilde\Omega$), hence $\gamma^{(k)} = \mathbb{E}(\gamma^{(k)})$ and $\Lambda_k = \mathbb{E}(\Lambda_k)$.

Proof. 1) Let

$$Y_n^k \;:=\; \log\|\wedge^k A_n\| \qquad (n\in\mathbb{N}, k = 1,\ldots,d)\,;$$

then $(Y_n^k)_n$ for each $k = 1,\ldots,d$ is subadditive: in case $k = 1$ this had been shown in 5.3; for $k > 1$ the calculation can immediately been transferred, since for all matrices B, C we have: $\wedge^k(BC) = (\wedge^k B)(\wedge^k C)$. Consequently with A also each $\wedge^k A$ is a *cocycle*, i.e. we have:

$$\wedge^k A_{n+m} \;=\; \wedge^k A_n\circ\varphi^m \,\cdot\, \wedge^k A_m\,.$$

Hence subadditivity of $(Y_n^k)_n$ follows.

2) The existence of $\tilde{\Omega}$ and $\gamma^{(k)}$ with the claimed properties follows from Theorem 5.7, applied to $(-Y_n^k)_n$; it remains to prove:

$$\gamma^{(k+m)} \leq \gamma^{(k)} + \gamma^{(m)} ;$$

but this follows directly from the characteristic property of the $\gamma^{(k)}$ and the norm inequality

$$\| \wedge^{k+m} A_n\| \leq \| \wedge^k A_n\| \cdot \| \wedge^m A_n\| .$$

3) We now prove the claims with respect to Λ_k : By Lemma 6.14 we have for $k = 1, \ldots, d$:

$$\frac{1}{n} \log \| \wedge^k A_n\| = \frac{1}{n} \sum_{i=1}^{k} \log \delta_i(A_n).$$

We have $\Lambda_1 \equiv \gamma^{(1)}$ and for $\omega \in \tilde{\Omega}$ we obtain successively:

$$\Lambda_{k+1}(\omega) \equiv \gamma^{(k+1)}(\omega) - \gamma^{(k)}(\omega) = \lim_{n\to\infty} \frac{1}{n} \log \delta_{k+1}(A_n) ,$$

if $\gamma^k(\omega) > \infty$; if this procedure ends, i.e. if $\gamma^{k_0}(\omega) = -\infty$, then also $\gamma^k(\omega) = -\infty$ for all $k \geq k_0$ and thus also $\Lambda_k = -\infty$ for all $k \geq k_0$. The remaining statements hold true by

$$\delta_1(A_n) \geq \delta_2(A_n) \geq \ldots \geq \delta_d(A_n)$$

and the respective expectations exist by hypothesis. □

7. The multiplicative ergodic theorem of Oseledets

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with a measure preserving mapping $\varphi : \Omega \to \Omega$ and $A : \Omega \to \mathbb{R}^{d\times d}$ a random matrix. According to (2) we define

$$A_n \;:=\; \begin{cases} \left(A\circ\varphi^{n-1}\right)\,\left(A\circ\varphi^{n-2}\right)\,\cdots\left(A\circ\varphi\right)\,A\,, & n\in\mathbb{N}\,,\\ I\,, & n=0\,,\end{cases}$$

the *cocycle generated by* A; A_n is therefore a *cocycle* over φ, i.e. we have:

$$A_{n+m} \;=\; (A_n\circ\varphi^m)\,\cdot\,A_m \qquad\qquad (m,n\in\mathbb{N}_0)\,,$$

as has already been used in the proof of the theorem of Furstenberg-Kesten.

We are now interested in the asymptotics of $|A_n x|$ for $x\in\mathbb{R}^d$ as $n\to\infty$. We trace this behavior back to the theorem of Furstenberg-Kesten by means of the following (deterministic) proposition 7.3. To prove the convergence statements therein we first establish two Lemmata:

Lemma 7.1. *Let* $\Phi\in\mathbb{R}^{d\times d}$ *by symmetric with spectral decomposition*

$$\Phi \;=\; \sum_{i=1}^{r}\lambda_i P_i\,,$$

where $r\le d$ *and* λ_i *the eigenvalues and* P_i *the corresponding orthogonal projectors on the eigenspaces. Let*

$$\Phi_n \;=\; \sum_{i=1}^{r_n}\lambda_i^n P_i^n$$

equally symmetric $d\times d$*-matrices, so that we have:*

i) $\lambda_k^n \xrightarrow{n\to\infty} \lambda_i$ *for all* $k\in\Sigma_i$ *, where* $\Sigma_i\neq\emptyset$ *are sets of indices (i=1,…, r);*

ii) $\bar{P}_i^n := \sum_{k\in\Sigma_i} P_k^n \xrightarrow{n\to\infty} P_i$ *for all* $i=1,\dots,r$ *.*

Then: $\Phi_n \xrightarrow{n\to\infty} \Phi$ *.*

Proof. With the convergence statements we obtain:

$$\begin{aligned}\Phi_n \;-\;\Phi \;&=\; \sum_{i=1}^{r}\sum_{k\in\Sigma_i}\lambda_k^n P_k^n \;-\; \sum_{i=1}^{r}\lambda_i P_i\\ &=\; \sum_{i=1}^{r}\Bigg[\underbrace{\sum_{k\in\Sigma_i}\left(\lambda_k^n-\lambda_i\right)\,P_k^n}_{\to 0} + \lambda_i\,\underbrace{\Big(\sum_{k\in\Sigma_i}P_k^n - P_i\Big)}_{\to 0}\Bigg] \;\xrightarrow{n\to\infty}\; 0\,.\end{aligned}$$

□

Lemma 7.2. *Let P, Q be orthogonal projectors in $\mathbb{R}^2$, such that we have:*

$$\dim U \;=\; \dim V \;=\; 1\,, \qquad \textit{where} \quad U := \operatorname{Im} P \textit{ and } V := \operatorname{Im} Q\,.$$

Then:

$$\delta(U,V) \;:=\; \|P-Q\| \;=\; |x\wedge y| \;=\; |\sin\alpha| \qquad (x\in U, y\in V \textit{ with } |x|=|y|=1)\,,$$

where α is the angle between x and y. Consequently δ is a complete metric on $\mathbb{P}^1$, the projective space of all one dimensional linear subspaces of $\mathbb{R}^2$.

Proof. The second equation has already been proved in the remark after Lemma 6.11.

$\|P-Q\| \;=\; |x\wedge y|$: As in the remark after Lemma 6.11 we further obtain:

$$\begin{aligned}
|x\wedge y| \;&=\; \det\begin{pmatrix} \langle x,x\rangle & \langle x,y\rangle \\ \langle y,x\rangle & \langle y,y\rangle \end{pmatrix}^{1/2} \\
&=\; \sqrt{1-\langle x,y\rangle^2} \\
&=\; \sqrt{\langle x,y\rangle^2+\langle x,y^\perp\rangle^2-\langle x,y\rangle^2} \\
&=\; |\,\langle x,y^\perp\rangle\,| \\
&=\; \|\,(I-Q)\,P\,\| \\
&=\; \|\,(P-Q)\,P\,\| \;\leq\; \|\,P-Q\,\|\,,
\end{aligned}$$

where the idempotence of orthogonal projectors has been used besides the fact $\|AB\|=\|BA\|$ for orthogonal projectors A, B.
Note that $||PQ||=|\langle x,y\rangle|=|\cos(\alpha)|=|\langle x^\perp,y^\perp\rangle|=||(I-Q)(I-P)||$.
On the other hand for $w\in\mathbb{R}^2$:

$$\begin{aligned}
|\,(P-Q)w\,|^2 \;&=\; |\,(P-QP)w-(Q-QP)w\,|^2 \\
&=\; |\,(I-Q)Pw-Q(I-P)w\,|^2 \\
&=\; |\,(I-Q)Pw\,|^2 \;+\; |\,Q(I-P)w\,|^2 \\
&\leq\; \|\,(I-Q)P\,\|^2\,|\,Pw\,|^2 \;+\; \underbrace{\|\,Q(I-P)\,\|^2}_{\|\,(I-Q)P\,\|^2}\;|\,(I-P)w\,|^2 \\
&=\; \|\,(I-Q)P\,\|^2\,,
\end{aligned}$$

hence

$$\|\,P-Q\,\| \;\leq\; \|\,(I-Q)P\,\|\,.$$

In summary we proved:

$$\|\,P-Q\,\| \;=\; \|\,(I-Q)P\,\| \;=\; |x\wedge y|\,.$$

□

The following deterministic theorem serves to prepare for an application of the theorem of Furstenberg-Kesten.

Proposition 7.3 (Goldsheid-Margulis). *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{d\times d}$ with the properties:*

$$\limsup_{n\to\infty} \frac{1}{n} \log \|A_n\| \;\leq\; 0 \tag{4}$$

and assume that $\Phi_n := A_n \cdots A_1$ *fulfils*

$$\lim_{n\to\infty} \frac{1}{n} \log \| \wedge^i \Phi_n \| \;=:\; \gamma^{(i)} \;\in\; \mathbb{R} \cup \{-\infty\} \tag{5}$$

for each $i = 1, \ldots, d$. *Then we have:*

i) There exists (in the topology of the operator norm) the limit

$$\Psi \;:=\; \lim_{n\to\infty} (\, \Phi_n^* \, \Phi_n \,)^{1/2n} \;\geq\; 0\,.$$

Defining successively Λ_i *for* $i = 1, \ldots, d$ *by* $\Lambda_1 + \cdots + \Lambda_i \;=\; \gamma^{(i)}$ *(if* $\gamma^{(i)} = -\infty$, *set* $\Lambda_i = -\infty$*), then the eigenvalues of* Ψ *are given by*

$$e^{\Lambda_1}, \ldots, e^{\Lambda_d}$$

and we have

$$\Lambda_i = \lim_{n\to\infty} \frac{1}{n} \log \delta_i(\Phi_n) \qquad\qquad (i = 1, \ldots, d)\,.$$

ii) Let

$$e^{\lambda_p} < \cdots < e^{\lambda_1}$$

the different (!) eigenvalues of Ψ *(where* $\lambda_p = -\infty$ *is possible),* $U_p, \ldots, U_1$ *the corresponding eigenspaces with* $d_i := \dim U_i$ *and let*

$$V_i \;:=\; \begin{cases} \{0\}\,, & i = p+1 \\ U_p \oplus \cdots \oplus U_i\,, & i = 1, \ldots, p\,. \end{cases}$$

Then we have:

$$V_{p+1} \subset V_p \subset V_{p-1} \subset \cdots \subset V_1 = \mathbb{R}^d$$

and for each $x \in \mathbb{R}^d \setminus \{0\}$ *there exists the* Lyapunov exponent

$$\lambda(x) \;:=\; \lim_{n\to\infty} \frac{1}{n} \log |\Phi_n x|\;;$$

we have for all $i = 1, \ldots, p$*:*

$$x \in V_i \setminus V_{i+1} \iff \lambda(x) = \lambda_i$$

resp. equivalently:

$$V_i \;=\; \{\, x \in \mathbb{R}^d \,:\, \lambda(x) \leq \lambda_i \,\}\,.$$

PROOF. In case $d = 1$ nothing needs to be proven, since then $\Phi_n \in \mathbb{R}$ and the claims follow directly from the hypotheses.
For simplicity we now confine our attention to the case $d = 2$; the general case can be proved in a similar way, with more technicalities (see Arnold [AR 98] pp. 144-152).

$\boxed{\Lambda_i = \lim_{n\to\infty} \frac{1}{n} \log \delta_i(\Phi_n) \text{ for } i = 1, 2}$ This follows from (5) with Lemma 6.14 iii):

$$\Lambda_1 \equiv \gamma^1 \overset{(5)}{=} \lim_{n\to\infty} \frac{1}{n} \log \|\Phi_n\| \overset{6.14}{=} \lim_{n\to\infty} \frac{1}{n} \log \delta_1(\Phi_n) \, ;$$

if now $\Lambda_1 = -\infty$, hence $\gamma^1 = -\infty$, by (5) also $\gamma^2 = -\infty = \Lambda_2$; on the other hand in this case

$$\frac{1}{n} \log \delta_2(\Phi_n) \le \frac{1}{n} \log \delta_1(\Phi_n) \longrightarrow -\infty.$$

If $\Lambda_1 > -\infty$, we get:

$$\begin{aligned} \Lambda_2 \equiv \gamma^2 - \Lambda_1 &= \lim_{n\to\infty} \frac{1}{n} \log \underbrace{\| \wedge^2 \Phi_n \|}_{\delta_1(\Phi_n)\delta_2(\Phi_n)} - \lim_{n\to\infty} \frac{1}{n} \log \delta_1(\Phi_n) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \delta_2(\Phi_n) \, . \end{aligned}$$

$\boxed{\text{convergence of operators and Lyapunov exponents}}$ Let now

$$\Phi_n = V_n D_n O_n$$

be the singular value decomposition of Φ_n, with

$$D_n = \begin{pmatrix} \delta_1(\Phi_n) & 0 \\ 0 & \delta_2(\Phi_n) \end{pmatrix} .$$

from this we obtain:

$$(\Phi_n^* \, \Phi_n)^{1/2n} = (O_n^* D_n^2 O_n)^{1/2n} = O_n^* D_n^{1/n} O_n \, ;$$

this matrix has eigenvalues $\delta_1(\Phi_n)^{1/n}$ and $\delta_2(\Phi_n)^{1/n}$, which according to what has been shown above converge to e^{Λ_1} and e^{Λ_2}; so we have the following convergences:

$$D_n^{1/n} \equiv \begin{pmatrix} \delta_1^{1/n}(\Phi_n) & 0 \\ 0 & \delta_2^{1/n}(\Phi_n) \end{pmatrix} \xrightarrow{n\to\infty} \begin{pmatrix} e^{\Lambda_1} & 0 \\ 0 & e^{\Lambda_2} \end{pmatrix} .$$

Now the difficulty is that the convergence of O_n in general is not guaranteed; but it is enough to prove convergence of the respective eigenspaces for which Lemma 7.1 has been established.

1. CASE: $\Lambda_1 = \Lambda_2 =: \lambda_1$: As just seen we have $D_n^{1/n} \to e^{\lambda_1} I$ and Lemma 7.1 imply

$$(\Phi_n^* \Phi_n)^{1/2n} \xrightarrow{n\to\infty} e^{\lambda_1} I$$

with $\bar{P}_1^n := P_1^n + P_2^n$. Moreover we have immediately: $V_1 \equiv U_1 = \mathbb{R}^2$, $p = 1$ and $d_1 = 2$. Therefore we only have to prove that for all $x \in \mathbb{R}^2 \setminus \{0\}$ we have:

$$\lambda(x) \;\equiv\; \lim_{n\to\infty} \frac{1}{n} \log \|\Phi_n x\| \;=\; \lambda_1 \,.$$

For this let first $\lambda_1 > -\infty$; then from the already proven characterization of Λ_1, it follows that for each $\epsilon > 0$ there exists $c_\epsilon \in (0, \infty)$ with

$$\frac{1}{c_\epsilon} e^{n(\lambda_1 - \epsilon)} \;\le\; \delta_i(\Phi_n) \;\le\; c_\epsilon e^{n(\lambda_1+\epsilon)}, \qquad i = 1, 2.$$

Setting $x_n := O_n x$, we get

$$|\Phi_n x| \;=\; |V_n D_n O_n x| \;=\; |D_n x_n| = \big(\delta_1(\Phi_n)^2 (x_n^1)^2 + \delta_2(\Phi_n)^2 (x_n^2)^2\big)^{1/2}$$

with x_n^i denoting the components of x_n; in summary we therefore have

$$\frac{|x|}{c_\epsilon}\, e^{n(\lambda_1-\epsilon)} \;\le\; |\Phi_n x| \;\le\; |x| c_\epsilon e^{n(\lambda_1+\epsilon)}\,,$$

whence we obtain that $\lambda(x) = \lambda_1$.
If $\lambda = -\infty$, we can in the same way find for each $r < 0$ a $c_r \in (0, \infty)$ such that

$$0 \;\le\; \delta_i(\Phi_n) \;\le\; c_r e^{nr}\,, \qquad i = 1, 2\,.$$

As above we then infer:

$$0 \;\le\; |\Phi_n x| \;\le\; |x| c_r e^{nr},$$

from which we conclude as above: $\lambda(x) = \lambda_1$. So the theorem is proved in case $\Lambda_1 = \Lambda_2$.

2. CASE: $\lambda_1 \equiv \Lambda_1 > \Lambda_2 \equiv \lambda_2$: Here we have

$$D_n^{1/n} \equiv \begin{pmatrix} \delta_1^{1/n}(\Phi_n) & 0 \\ 0 & \delta_2^{1/n}(\Phi_n) \end{pmatrix} \xrightarrow{n\to\infty} \begin{pmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{pmatrix}.$$

To prove the existence of Ψ, we have to show that the orthogonal projectors P_1^n, P_2^n on the eigenspaces U_1^n, U_2^n of $(\Phi_n^*\Phi_n)^{1/2n}$ converge to orthogonal projectors P_1, P_2, since then Lemma 7.1 yields:

$$(\Phi_n^*\, \Phi_n)^{1/2n} \;\xrightarrow{n\to\infty}\; e^{\lambda_1} P_1 + e^{\lambda_2} P_2 \;=:\; \Psi\,.$$

This will be proved in the following Lemma by means of a Cauchy sequence argument. For this purpose we remark that the eigenvectors of $(\Phi_n^*\Phi_n)^{1/2n} = O_n^* D_n^{1/n} O_n$ are given by $u_i^n := O_n^* e_i$ $(i = 1, 2)$, where (e_1, e_2) is the standard basis of $\mathbb{R}^2$. In particular $U_i^n = \text{span}\,(u_i^n)$, $i = 1, 2$.

Lemma 7.4. *In the situation above ("‘2. case"’ in the proof of theorem 7.3) we have:*

$$\limsup_{n\to\infty} \frac{1}{n} \log \delta\big(U_i^n, U_i^{n+1}\big) \;\le\; \lambda_2 - \lambda_1 \;<\; 0 \qquad (i = 1, 2).$$

In particular $(U_i^n)_{n\in\mathbb{N}}$ *($i = 1, 2$) is a Cauchy sequence in the projective space* P^1*, that hence converges to* $U_i \in P^1$*. Moreover, this convergence takes place with exponential speed:*

$$\limsup_{n\to\infty} \frac{1}{n} \log \delta\big(U_i^n, U_i\big) \;\le\; \lambda_2 - \lambda_1 \qquad (i = 1, 2).$$

Proof of Lemma 7.4. Wlog let hereby $i = 2$, since U_1^n is orthogonal to U_2^n, but the metric δ on $\mathbb{P}^1$ is invariant with respect to orthogonal transformations. By orthogonality of all (u_1^{n+1}, u_2^{n+1}) we may represent u_2^n as

$$u_2^n \;=\; \alpha_n u_1^{n+1} + \beta_n u_2^{n+1} \qquad (n \in \mathbb{N})\,.$$

1) $\delta\big(U_2^n, U_2^{n+1}\big) = |\alpha_n|$, because:

$$\begin{aligned}\delta\big(U_2^n, U_2^{n+1}\big) \overset{7.2}{=} |u_2^n \wedge u_2^{n+1}| \;&\equiv\; |(\alpha_n u_1^{n+1} + \beta_n u_2^{n+1}) \wedge u_2^{n+1}|\\ &=\; |\alpha_n|\, |u_1^{n+1} \wedge u_2^{n+1}|\\ &=\; |\alpha_n|\,,\end{aligned}$$

where orthonormality of u_1^n and u_2^n was used.

2) $\delta\big(U_2^n, U_2^{n+1}\big) \le \|A_{n+1}\| \frac{\delta_2(\Phi_n)}{\delta_1(\Phi_{n+1})}$, because: First we have

$$\begin{aligned}|\Phi_{n+1} u_2^n| \;&\equiv\; |\alpha_n \Phi_{n+1} u_1^{n+1} + \beta_n \Phi_{n+1} u_2^{n+1}|\\ &\equiv\; |\alpha_n\, V_{n+1} D_{n+1} O_{n+1}\, O_{n+1}^* e_1 \;+\; \beta_n\, V_{n+1} D_{n+1} O_{n+1}\, O_{n+1}^* e_2|\\ &=\; |\alpha_n\, \delta_1(\Phi_{n+1})\, V_{n+1} e_1 \;+\; \beta_n\, \delta_2(\Phi_{n+1})\, V_{n+1} e_2|\\ &\overset{\text{orth.}}{\ge}\; |\alpha_n\, \delta_1(\Phi_{n+1})\, V_{n+1} e_1|\\ &=\; |\alpha_n|\, \delta_1(\Phi_{n+1})\,;\end{aligned}$$

on the other hand

$$|\Phi_{n+1} u_2^n| \;\equiv\; |A_{n+1} \Phi_n u_2^n| \;\le\; \|A_{n+1}\|\, |\Phi_n u_2^n| \;=\; \|A_{n+1}\|\, \delta_2(\Phi_n)\,,$$

hence in summary

$$\delta\big(U_2^n, U_2^{n+1}\big) \overset{1)}{=} |\alpha_n| \;\le\; \frac{|\Phi_{n+1} u_2^n|}{\delta_1(\Phi_{n+1})} \;\le\; \|A_{n+1}\|\, \frac{\delta_2(\Phi_n)}{\delta_1(\Phi_{n+1})}\,.$$

3) First claim of the Lemma: by what has just been proved we get:

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\delta\big(U_2^n,U_2^{n+1}\big) &\le \limsup_{n\to\infty}\frac{1}{n}\log\|A_{n+1}\| \\ &\quad+\limsup_{n\to\infty}\frac{1}{n}\log\delta_2(\Phi_n) \\ &\quad-\liminf_{n\to\infty}\frac{1}{n}\log\delta_1(\Phi_{n+1}) \\ &\le 0+\lambda_2-\lambda_1\,,\end{aligned}$$

where we use the first hypothesis of theorem 7.3 and the convergence result already established.

4) $(U_2^n)_n$ converges in P^1 to some U_2: For this purpose let $\varepsilon<\lambda_1-\lambda_2$; by what has been shown we can choose $n_0\in\mathbb{N}$ such that

$$\frac{1}{n}\log\delta\big(U_2^n,U_2^{n+1}\big)\;<\;\lambda_2-\lambda_1+\varepsilon\quad(<0)\qquad(\forall\, n\ge n_0).$$

But then we get for $n_0\le m\le n$:

$$\begin{aligned}\delta\big(U_2^n,U_2^{n+1}\big) &\le \sum_{k=m}^{n-1}\delta\big(U_2^k,U_2^{k+1}\big) \\ &\le \sum_{k=m}^{n-1}\mathrm{e}^{k(\lambda_2-\lambda_1+\varepsilon)} \\ &\le \sum_{k=m}^{\infty}\mathrm{e}^{k(\lambda_2-\lambda_1+\varepsilon)} \\ &= \frac{\mathrm{e}^{m(\lambda_2-\lambda_1+\varepsilon)}}{1-\mathrm{e}^{\lambda_2-\lambda_1+\varepsilon}}\;\xrightarrow{m\to\infty}\;0\,,\end{aligned}$$

where the summation formula for geometric series was employed.

5) Second claim of the Lemma: with the arguments just used we also get:

$$\delta\big(U_2^n,U_2\big)\;\le\;\mathrm{e}^{n(\lambda_2-\lambda_1+\varepsilon)}\,\frac{1}{1-\mathrm{e}^{\lambda_2-\lambda_1+\varepsilon}}$$

and therefore

$$\limsup_{n\to\infty}\frac{1}{n}\log\delta\big(U_2^n,U_2\big)\;\le\;\lambda_2-\lambda_1+\varepsilon\,.$$

Now the claim follows with $\varepsilon\to 0$. □

Lem. 7.4

Continuation of the proof of Proposition 7.3. As orthogonal projectors P_1, P_2 we now choose the projectors on the spaces U_1, U_2 existing due to Lemma 7.4. By Lemma 7.2 and Lemma 7.4 we get the convergence

$$P_i^n \xrightarrow{n\to\infty} P_i \qquad (i = 1, 2)\,.$$

In summary we obtain

$$(\Phi_n^* \, \Phi_n)^{1/2n} \xrightarrow{n\to\infty} e^{\lambda_1} P_1 + e^{\lambda_2} P_2 \ =: \ \Psi$$

It remains to prove the claim on the *Lyapunov exponents*; hereby $V_2 = U_2 \subset \mathbb{R}^2 = V_1$, such that it remains to prove:

$$x \in V_2 \setminus \{0\} \implies \lim_{n\to\infty} \frac{1}{n} \log \|\Phi_n x\| = \lambda_2 \qquad \text{and}$$
$$x \in \mathbb{R}^2 \setminus V_2 \implies \lim_{n\to\infty} \frac{1}{n} \log \|\Phi_n x\| = \lambda_1\,;$$

where in each case we may assume $|x| = 1$.

$x \in V_2 \setminus \{0\} \Rightarrow \lim \frac{1}{n} \log |\Phi_n x| = \lambda_2$: We represent x as

$$x \ = \ \alpha_n u_1^n + \beta_n u_2^n\ ,$$

hence again

$$\Phi_n x \ = \ \alpha_n \Phi_n u_1^n + \beta_n \Phi_n u_2^n \ = \ \alpha_n\, \delta_1(\Phi_n)\, V_n e_1 + \beta_n\, \delta_2(\Phi_n)\, V_n e_2\ ,$$

and thus

$$|\beta_n|\, \delta_2(\Phi_n) \ \le \ \left[\alpha_n^2\, \delta_1(\Phi_n)^2 + \beta_n^2\, \delta_2(\Phi_n)^2\right]^{1/2} \ = \ |\Phi_n x|\ ;$$

as in the proof of 1) of Lemma 7.4 we obtain from Lemma 7.2: $\delta\big(U_2^n, U_2^{n+1}\big) = |\alpha_n|$, since $x \in V_2 = U_2$; consequently by Lemma 7.4 also

$$\limsup_{n\to\infty} \frac{1}{n} \log |\alpha_n| \ = \ \limsup_{n\to\infty} \frac{1}{n} \log \delta\big(U_2^n, U_2\big) \ \le \ \lambda_2 - \lambda_1 \ < \ 0\,;$$

thus we infer:

$$\beta_n^2 \ = \ 1 - \alpha_n^2 \xrightarrow{n\to\infty} 1\,.$$

and therefore in summary:

$$\begin{aligned}
\lambda_2 \ &= \ \lim_{n\to\infty} \frac{1}{n} \log\left(\,|\beta_n|\, \delta_2(\Phi_n)\,\right) \\
&\le \ \liminf_{n\to\infty} \frac{1}{n} \log |\Phi_n x| \ \le \ \limsup_{n\to\infty} \frac{1}{n} \log |\Phi_n x| \\
&= \ \frac{1}{2}\, \limsup_{n\to\infty} \frac{1}{n} \log\left[\,\alpha_n^2\, \delta_1(\Phi_n)^2 + \beta_n^2\, \delta_2(\Phi_n)^2\right] \\
&\le \ \frac{1}{2}\, \max\left\{ \limsup_{n\to\infty} \frac{1}{n} \log \alpha_n^2\, \delta_1(\Phi_n)^2 \ , \ \limsup_{n\to\infty} \frac{1}{n} \log \beta_n^2\, \delta_2(\Phi_n)^2 \right\} \\
&\le \ \max\left\{\,(\lambda_2 - \lambda_1) + \lambda_1 \ , \ 0 + \lambda_2\,\right\} \\
&= \ \lambda_2\,.
\end{aligned}$$

$\underline{x \in \mathbb{R}^2 \setminus V_2 \ \Rightarrow \lim \frac{1}{n} \log |\Phi_n x| = \lambda_1}$: Here we represent x as

$$x \ = \ \alpha u + \beta v$$

with unit vectors $u \in U_1$ and $v \in U_2 = V_2$; the latter are represented by

$$v \ = \ \alpha_n u_1^n + \beta_n u_2^n \qquad \text{resp.} \qquad u \ = \ \gamma_n u_1^n + \delta_n u_2^n \ .$$

Also in this case we necessarily get from Lemma 7.4: $\alpha_n \to 0$, $\delta_n \to 0$ and thus $|\beta_n| \to 1$, $|\gamma_n| \to 1$ (in projective space we have by Lemma 7.4: $u_1^n \to u$ and $u_2^n \to v$).
Therefore we have as above:

$$\begin{aligned} |\alpha|\,|\gamma_n|\,\delta_1(\Phi_n) \ &\leq \ \left[\,(\alpha\gamma_n + \beta\alpha_n)^2\,\delta_1(\Phi_n)^2 + (\alpha\delta_n + \beta\beta_n)^2\,\delta_2(\Phi_n)^2\right]^{1/2} \\ &= \ |\Phi_n x| \ ; \end{aligned}$$

noting that by the position of x always $\alpha = \langle x, u\rangle \neq 0$, we in summary again obtain:

$$\begin{aligned} \lambda_1 \ &= \ \lim_{n\to\infty} \frac{1}{n} \log\left(\,|\alpha|\,|\gamma_n|\,\delta_1(\Phi_n)\,\right) \\ &\leq \ \liminf_{n\to\infty} \frac{1}{n} \log|\Phi_n x| \ \leq \ \limsup_{n\to\infty} \frac{1}{n} \log|\Phi_n x| \\ &= \ \frac{1}{2} \limsup_{n\to\infty} \frac{1}{n} \log\left[\,(\alpha\gamma_n + \beta\alpha_n)^2\,\delta_1(\Phi_n)^2 + (\alpha\delta_n + \beta\beta_n)^2\,\delta_2(\Phi_n)^2\right] \\ &\leq \ \lambda_1 \ . \end{aligned}$$

Thus all claims of Proposition 7.3 have been proven. □

To be able to apply the Proposition by Goldsheid-Margulis, it remains to check the first hypothesis in the special case of stationary random matrices:

Lemma 7.5. *Let $X : \Omega \to \mathbb{R} \cup \{-\infty\}$ be a random variable with $X^+ \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Then*

$$\Omega_1 \ := \ \left\{ \limsup_{n\to\infty} \frac{1}{n} X \circ \varphi^{n-1} \ \leq \ 0 \right\}$$

is invariant and we have $\mathbb{P}(\Omega_1) = 1$.

Proof. The invariance follows from the definition of Ω_1. Moreover Ω_1 has full measure, because:

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left\{\frac{1}{n} X \circ \varphi^{n-1} \ > \ \varepsilon\right\} \ &\overset{\varphi \text{ m. pres.}}{=} \ \sum_{n=1}^{\infty} \mathbb{P}\{X > \varepsilon n\} \ = \ \sum_{n=1}^{\infty} \mathbb{P}\left\{X^+ > \varepsilon n\right\} \\ &\leq \ \frac{1}{\varepsilon}\,\mathbb{E}(X^+) \ < \ \infty \, , \end{aligned}$$

hence by Borel-Cantelli: $\mathbb{P}(\Omega_1) = 1$. □

To deduce the main theorem, we apply Lemma 7.5 to $X := \log \|A\|$. So we obtain:

Theorem 7.6 (Multiplicative ergodic theorem, Oseledets)**.** *Let* $A : \Omega \to \mathbb{R}^{d\times d}$ *be a random matrix on* $(\Omega, \mathscr{F}, \mathbb{P}, \varphi)$ *and*

$$A_n := \begin{cases} \left(A\circ\varphi^{n-1}\right)\left(A\circ\varphi^{n-2}\right)\cdots(A\circ\varphi)\,A\,, & n\in\mathbb{N}\,,\\ I\,, & n=0\,,\end{cases}$$

the cocycle on $\mathbb{R}^d$ *generated by this sequence. Assume*

$$\log^+\|A\| \in L^1(\Omega,\mathscr{F},\mathbb{P})\,.$$

Then there exists $\tilde\Omega \in \mathscr{F}$ *with* $\tilde\Omega \subset \varphi^{-1}(\tilde\Omega)$ *and* $\mathbb{P}(\tilde\Omega)=1$, *such that for each* $\omega\in\tilde\Omega$ *we have:*

i) There exists

$$\Psi(\omega) := \lim_{n\to\infty}\left(A_n^*(\omega)\,A_n(\omega)\right)^{1/2n} \ \ge\ 0.$$

ii) If

$$e^{\lambda_{p(\omega)}(\omega)} < \cdots < e^{\lambda_1(\omega)}$$

are the different eigenvalues of $\Psi(\omega)$ *(where* $\lambda_{p(\omega)}(\omega) = -\infty$ *is possible), and* $U_{p(\omega)}(\omega),\dots,U_1(\omega)$ *are the corresponding eigenspaces with* $d_i(\omega) := \dim U_i(\omega)$, *then we have:*

$$\big(\lambda_i\circ\varphi\big)(\omega) = \lambda_i(\omega)\,,\quad \big(d_i\circ\varphi\big)(\omega) = d_i(\omega)\,,\quad \text{and}\quad 1\le i\le p_i(\omega) = \big(p_i\circ\varphi\big)(\omega)\,.$$

iii) Defining

$$V_i(\omega) := \begin{cases}\{0\}\,, & i = p(\omega)+1\\ U_{p(\omega)}(\omega)\oplus\cdots\oplus U_i(\omega)\,, & i=1,\dots,p(\omega)\,,\end{cases}$$

we have

$$V_{p(\omega)+1}(\omega)\subset V_{p(\omega)}(\omega)\subset V_{p(\omega)-1}(\omega)\subset\cdots\subset V_1(\omega) = \mathbb{R}^d$$

and for each $x\in\mathbb{R}^d\setminus\{0\}$ *there exists*

$$\lambda(\omega,x) := \lim_{n\to\infty}\frac{1}{n}\log|A_n(\omega)x|\,;$$

and we have for all $i=1,\dots,p(\omega)$*:*

$$x\in V_i(\omega)\setminus V_{i+1}(\omega) \iff \lambda(\omega,x) = \lambda_i(\omega)$$

resp. equivalently:

$$V_i(\omega) = \{\,x\in\mathbb{R}^d : \lambda(\omega,x)\le\lambda_i(\omega)\,\}\,.$$

iv) If φ *is ergodic, then* p,λ_i *and* d_i *on* $\tilde\Omega$ *are constant* $\mathbb{P}$*-a.s..*

Proof. By the integrability hypothesis Lemma 7.5 is applicable with $X := \log \|A\|$ and provides the invariant set

$$\tilde{\Omega}_1 := \left\{ \omega \in \Omega : \limsup_{n\to\infty} \frac{1}{n} \log \|A(\varphi^{n-1}\omega)\| \le 0 \right\}$$

with full measure. We now apply the deterministic MET Proposition 7.3 to

$$A_n^\omega := A(\varphi^{n-1}\omega) \qquad \text{and} \qquad \Phi_n^\omega \equiv A_n^\omega \cdots A_1^\omega \overset{\text{cocycle}}{=} A_n(\omega)\,,$$

where (4) is valid by definition on $\tilde{\Omega}_1$ and (5) holds true by the theorem of Furstenberg-Kesten 6.15 on a forward invariant set $\tilde{\Omega}_2$ with full measure; consequently Proposition 7.3 is applicable for each $\omega \in \tilde{\Omega}_1 \cap \tilde{\Omega}_2 =: \tilde{\Omega}$, a forward invariant set of full measure, and yields with Theorem 6.15 the desired claims. □

Definition 7.7. The functions λ_i from the theorem of Oseledets are called *Lyapunov exponents* of the linear cocycle $(A_n)_{n\in\mathbb{N}_0}$.

The spaces V_i (for $i = 1, \ldots, p$) are not the analogues of eigenspaces from the deterministic theory. For such an analogy the theory has to be extended to cocycles indexed by $\mathbb{Z}$, see Arnold [AR 98], Theorem 3.4.11. .

Notations

$\mathbb{R}_+$	$\{t \in \mathbb{R} : t \geq 0\}$
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$s^\pm$	$(\pm s) \vee 0$; positive resp. negative part of a real number or function s
$\equiv$	equality by definition
$\vert\ \vert$	norm
$\Vert\ \Vert$	operator norm
$M \dot\cup N$	disjoint union of M and N
$\mathscr{B}(X)$	Borel σ-algebra on the topological space X
$\mathscr{B}^n$	$\mathscr{B}(\mathbb{R}^n)$
$\delta_1(A) \geq \cdots \geq \delta_d(A)$	singular values of $A \in \mathbb{R}^{d \times d}$
$\mathbb{E}(f)$	$\int f \, d\mathbb{P}$; expectation of a function f with respect to the probability measure $\mathbb{P}$
$\mathbb{E}(f \vert \mathscr{F})$	conditional expectation of the random variable f given $\mathscr{F}$
$\mathscr{I}$	σ-algebra of measurable invariant sets
$\sigma(\mathscr{M})$	σ-Algebra generated by a family $\mathscr{M}$ of sets resp. functions
RV	random variable
Wlog	without loss of generality

References

[AR 98] L. Arnold. *Random Dynamical Systems.* Springer-Verlag, Berlin Heidelberg, 1998.

[G-M 89] I.Y. Goldsheid & G.A. Margulis. Lyapunov Indices of a product of random matrices. *Russian Mathematical Surveys*, 44:11-71, 1989.

[HM 74] P. Halmos. *Measure Theory.* Springer-Verlag, New York, 1974.

[KB 97] O. Kallenberg. *Foundations of Modern Probability.* Springer-Verlag, New York, 1997.

[KI 68] J.F.C. Kingman. The ergodic theory of subadditive stochastic processes. *J. Royal Statist. Soc. Ser. B*, 30:499-510, 1968.

[M-T 93] S. Meyn & R. Tweedie. *Markov chains and stochastic stability* . Springer, London, 1993.

[OS 68] V.I. Oseledets. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19:197-231, 1968.

[Sh 95] A.N. Shiryaev. *Probability.* (second edition) GTM 95 Springer-Verlag, Berlin Heidelberg NewYork, 1995.

Index