ANNALES
POLONICI MATHEMATICI
LXXIII.2 (2000)

# Generic properties of learning systems

by Tomasz Szarek (Katowice)

**Abstract.** It is shown that the set of learning systems having a singular stationary distribution is generic in the family of all systems satisfying the average contractivity condition.

**0. Introduction.** Generic properties of Markov operators have been studied in [1, 4–6, 8]. Lasota and Myjak [6] have proved that for most nonexpansive iterated function systems the corresponding stationary distribution is in fact singular.

In this paper we investigate iterated function systems with place dependent probabilities, so called learning systems. We prove results analogous to those obtained by Lasota and Myjak. Namely, in the family of all learning systems $(S,p)$ satisfying the condition $\lambda_{(S,p)} = \max_{x\in X}\sum_{i=1}^{N} p_i(x)L_i \le 1$, most have a singular stationary distribution. The proof of the main theorem is similar in spirit to the proof of Lasota and Myjak. The important difference is that the iterated function systems we study are not nonexpansive in the same norm. In the case examined by Lasota and Myjak all iterated function systems are nonexpansive in the Hutchinson norm (see [6]).

The organization of the paper is as follows. In Section 1 we introduce definitions and notation. Section 2 contains auxiliary lemmas and theorems which are used in proving the main result of the paper. The main theorem is proved in Section 3.

**1. Definitions and notation.** Let $X \subset \mathbb{R}^k$ be a compact convex set of positive (Lebesgue) measure $m(X)$. Let $\mathcal{B}(X)$ denote the $\sigma$-algebra of Borel subsets of $X$ and let $\mathcal{M}$ denote the family of all finite Borel measures on $\mathcal{B}(X)$. We denote by $\mathcal{M}_1$ the set of all $\mu \in \mathcal{M}$ such that $\mu(X) = 1$. The elements of $\mathcal{M}_1$ are called *distributions*.

2000 *Mathematics Subject Classification*: Primary 60J05, 28A80; Secondary 47A35, 58F08.

*Key words and phrases*: Markov operators, iterated function systems.

A measure $\mu \in \mathcal{M}$ is called *absolutely continuous* if $\mu(A) = 0$ for every $A \in \mathcal{B}(X)$ such that $m(A) = 0$, and it is called *singular* if there is $Y \in \mathcal{B}(X)$ with $m(Y) = 0$ such that $\mu(Y) = \mu(X)$. By the Lebesgue Decomposition Theorem every measure $\mu \in \mathcal{M}$ can be written in the form $\mu = \mu_{\rm a} + \mu_{\rm s}$, where $\mu_{\rm a}$ is absolutely continuous and $\mu_{\rm s}$ is singular.

Let

$$\mathcal{M}_{\rm sig} = \{\mu_1 - \mu_2 : \mu_1, \mu_2 \in \mathcal{M}\}$$

be the space of all signed finite (Borel) measures on $X$. For every $l \ge 1$ we introduce the *Fortet–Mourier norm* (see [2, 3])

$$\|\nu\|_l = \sup\{|\langle f, \nu\rangle| : f \in F_l\},$$

where $\langle f, \nu\rangle = \int_X f(x)\,\nu(dx)$ and $F_l$ is the space of all continuous functions $f : X \to \mathbb{R}$ such that $\sup_{x\in X} |f(x)| \le 1$ and $|f(x) - f(y)| \le l\|x - y\|$ (here $\|\cdot\|$ denotes a norm in $\mathbb{R}^k$).

It can be proved (see [2]) that the convergence

$$\lim_{n\to\infty} \|\mu_n - \mu\|_l = 0 \quad \text{ for } \mu_n, \mu \in \mathcal{M}_1,\ l \ge 1$$

is equivalent to the condition

$$\lim_{n\to\infty} \langle f, \mu_n\rangle = \langle f, \mu\rangle \quad \text{ for } f \in C(X),$$

i.e. to the weak convergence of the sequence $(\mu_n)_{n\ge1}$ to $\mu$ (here $C(X)$ stands for the space of all continuous functions $f : X \to \mathbb{R}$). Hence the norms $\|\cdot\|_{l_1}$ and $\|\cdot\|_{l_2}$ for $l_1, l_2 \ge 1$ are equivalent.

An operator $P : \mathcal{M} \to \mathcal{M}$ is called a *Markov operator* if it satisfies the following two conditions:

(i) *positive linearity*:

$$P(\lambda_1\mu_1 + \lambda_2\mu_2) = \lambda_1 P\mu_1 + \lambda_2 P\mu_2$$

for $\lambda_1, \lambda_2 \ge 0$ and $\mu_1, \mu_2 \in \mathcal{M}$,

(ii) *preservation of the norm*:

$$P\mu(X) = \mu(X) \quad \text{ for } \mu \in \mathcal{M}.$$

A Markov operator $P : \mathcal{M} \to \mathcal{M}$ is called *nonexpansive in the norm* $\|\cdot\|_l$, $l \ge 1$, if

$$\|P\mu_1 - P\mu_2\|_l \le \|\mu_1 - \mu_2\|_l \quad \text{ for } \mu_1, \mu_2 \in \mathcal{M}_1.$$

A measure $\mu \in \mathcal{M}$ is called *stationary* or *invariant* if $P\mu = \mu$. A Markov operator $P$ is called *asymptotically stable* if there exists a stationary distribution $\mu_\star$ such that

$$\lim_{n\to\infty} \langle f, P^n\mu\rangle = \langle f, \mu_\star\rangle \quad \text{ for } \mu \in \mathcal{M}_1,\ f \in C(X).$$

Fix an integer $N \ge 1$. By a *learning system*

$$(S, p) = (S_1, \ldots, S_N, p_1, \ldots, p_N)$$

we mean a finite sequence of continuous transformations $S_i : X \to X$ and continuous functions $p_i : X \to [0,1], i = 1, \ldots, N$, such that $\sum_{i=1}^N p_i(x) = 1$. The sequence $(p_i)_{i=1}^N$ as above is called a *probability vector*. We assume that $S_i$ is Lipschitzian with Lipschitz constant $L_i$ for $i = 1, \ldots, N$.

For a learning system $(S, p)$ the value

$$\lambda_{(S,p)} = \max_{x \in X} \sum_{i=1}^N p_i(x) L_i$$

plays an important role.

We denote by $\mathcal{F}$ the set of all learning systems $(S, p)$ such that $\lambda_{(S,p)} \le 1$. In $\mathcal{F}$ we introduce a metric $d$ defined by

$$d((S,p),(T,q)) = \sum_{i=1}^N \max_{x \in X} |p_i(x) - q_i(x)| + \sum_{i=1}^N \max_{x \in X} \|S_i(x) - T_i(x)\|$$

for $(S,p),(T,q) \in \mathcal{F}$. It is easy to prove that $\mathcal{F}$ endowed with the metric $d$ is a complete metric space.

For a given learning system $(S, p)$ we define the corresponding Markov operator $P_{(S,p)} : \mathcal{M} \to \mathcal{M}$ by

$$P_{(S,p)}\mu(A) = \sum_{i=1}^N \int\limits_{S_i^{-1}(A)} p_i(x)\, \mu(dx) \quad \text{ for } A \in \mathcal{B}(X)$$

and its adjoint $U_{(S,p)} : C(X) \to C(X)$ by

$$U_{(S,p)} f(x) = \sum_{i=1}^N p_i(x) f(S_i(x)).$$

We say that the learning system $(S, p)$ has a *stationary distribution* (resp. is *asymptotically stable*) if the corresponding Markov operator $P_{(S,p)}$ has a stationary distribution (resp. is asymptotically stable).

Finally recall that a subset of a complete metric space $\mathcal{X}$ is called *residual* if its complement is a set of first Baire category.

**2. Preliminary results.** In this section we recall some auxiliary facts and prove an easy lemma.

From Lemmas 2.1 and 2.2 of [6] one can deduce the following lemma:

LEMMA 2.1. *If a learning system $(S, p)$ has a unique stationary distribution $\mu_*$, then either $\mu_*$ is absolutely continuous or $\mu_*$ is singular.*

LEMMA 2.2. *Let $S : X \to X$ be a Lipschitz transformation with Lipschitz constant $L_S$. Then for every $\varepsilon > 0$ and $x_0 \in X$ there exists $r > 0$ and a Lipschitz transformation $T : X \to X$ with Lipschitz constant $L_T$ such that*:

(i) $L_T < L_S + \varepsilon$,
(ii) $\max_{x\in X} \|S(x) - T(x)\| < \varepsilon$,
(iii) $T(x) = S(x_0)$ *for* $\|x - x_0\| \le r$.

The proof can be found in [6, Lemma 3.5].

LEMMA 2.3. *Let* $(p_1, \ldots, p_N)$ *be a probability vector such that* $p_i : X \to [0,1]$ *are continuous. Then for every* $\varepsilon > 0$ *there exists a probability vector* $(q_1, \ldots, q_N)$ *such that* $q_i : X \to [0,1]$ *are Lipschitzian and*

$$q_i(x) > 0 \quad \text{and} \quad |p_i(x) - q_i(x)| < \varepsilon \tag{2.1}$$

*for* $i = 1, \ldots, N$ *and* $x \in X$.

Proof. Fix $\varepsilon > 0$. Choose $\delta > 0$ such that $4\delta N < \varepsilon$. By the Stone Theorem we find a sequence $(\overline{r}_1, \ldots, \overline{r}_N)$ of Lipschitzian functions satisfying

$$\max_{x\in X} |p_i(x) - \overline{r}_i(x)| < \delta \quad \text{for } i = 1, \ldots, N.$$

We set $r_i(x) = \max(\overline{r}_i(x), \delta)$ and

$$q_i(x) = \frac{r_i(x)}{\sum_{j=1}^N r_j(x)} \quad \text{for } x \in X,\ i = 1, \ldots, N.$$

Then

$$\begin{aligned}\sum_{i=1}^N |p_i(x) - r_i(x)| &\le \sum_{i=1}^N |p_i(x) - \overline{r}_i(x)| + \sum_{i=1}^N |\overline{r}_i(x) - r_i(x)| \\ &\le \delta N + \delta N = 2\delta N\end{aligned}$$

and consequently

$$\Big|1 - \sum_{i=1}^N r_i(x)\Big| = \Big|\sum_{i=1}^N p_i(x) - \sum_{i=1}^N r_i(x)\Big| \le \sum_{i=1}^N |p_i(x) - r_i(x)| \le 2\delta N.$$

Hence

$$\begin{aligned}\sum_{i=1}^N |p_i(x) - q_i(x)| &\le \sum_{i=1}^N |p_i(x) - r_i(x)| + \sum_{i=1}^N \Big| r_i(x) - r_i(x)\Big(\sum_{j=1}^N r_j(x)\Big)^{-1}\Big| \\ &\le 2\delta N + \sum_{i=1}^N r_i(x)\Big|1 - \Big(\sum_{j=1}^N r_j(x)\Big)^{-1}\Big| \\ &\le 2\delta N + \Big|\sum_{i=1}^N r_i(x) - 1\Big| \le 4\delta N < \varepsilon.\end{aligned}$$

Since the $r_i$ are Lipschitzian and $X$ is compact, the functions $q_i$ are Lipschitzian. ■

Let $\mathcal{F}_0$ be the set of all $(S, p) \in \mathcal{F}$ with the following properties:

(2.2) $\lambda_{(S,p)} < 1$,
(2.3) $p_i$ is Lipschitzian and $p_i(x) > 0$ for $i \in \{1, \ldots, N\}$ and $x \in X$,
(2.4) the stationary distribution $\mu_{(S,p)}$ corresponding to $(S,p)$ is singular.

We are now in a position to recall the following theorem.

PROPOSITION 2.1. *Let $P$ be a Markov operator nonexpansive in the norm $\|\cdot\|_l$ for some $l \geq 1$. Assume that for every $\varepsilon > 0$ there is a Borel set $A$ with* $\operatorname{diam} A \leq \varepsilon$, *a real number $\alpha > 0$ and an integer $\overline{n}$ such that*

$$P^{\overline{n}}\mu(A) \geq \alpha \quad \text{for } \mu \in \mathcal{M}_1. \tag{2.6}$$

*Then $P$ is asymptotically stable and for every $\varepsilon > 0$ there exists an integer $n$ such that*

$$\|P^n\mu_1 - P^n\mu_2\|_l < \varepsilon \quad \text{for } \mu_1, \mu_2 \in \mathcal{M}_1. \tag{2.7}$$

For details see the proof of Theorem 3.1 in [9]. In fact Theorem 3.1 was proved for $l = 1$ but the same argument works for every $l \geq 1$.

REMARK 2.1. If a learning system $(S,p) \in \mathcal{F}$ satisfies conditions (2.2) and (2.3), then the assumptions of Proposition 2.1 are satisfied. For details see the proof of Theorem 4.2 in [10].

## 3. Generic singularity of learning systems

LEMMA 3.1. *The set $\mathcal{F}_0$ is dense in the space $(\mathcal{F}, d)$.*

Proof. Fix $(S,p) \in \mathcal{F}$ and $\varepsilon > 0$. Let $z \in X$. Since $X$ is convex we can define for $i \in \{1, \ldots, N\}$ new transformations $\overline{S}_i : X \to X$ by

$$\overline{S}_i(x) = \alpha z + (1-\alpha)S_i(x) \quad \text{for } x \in X,$$

where $\alpha = \varepsilon(4N \operatorname{diam} X)^{-1}$. It follows immediately that

$$d((\overline{S},p),(S,p)) \leq \varepsilon/4 \tag{3.1}$$

and

$$\lambda_{(\overline{S},p)} \leq 1 - \alpha. \tag{3.2}$$

Thus there exists $i_0 \in \{1, \ldots, N\}$ such that $\overline{L}_{i_0} := \operatorname{Lip} \overline{S}_{i_0} < 1$. Let $x_0 \in X$ be a fixed point of $\overline{S}_{i_0}$. By Lemma 2.2 we find $\overline{T} : X \to X$ with Lipschitz constant $L_{\overline{T}}$ and with the following properties:

$$L_{\overline{T}} < \overline{L}_{i_0} + \eta, \tag{3.3}$$

$$\max_{x \in X} \|\overline{T}(x) - \overline{S}_{i_0}(x)\| < \varepsilon/4, \tag{3.4}$$

$$\overline{T}(x) = x_0 \quad \text{for } \|x - x_0\| \leq r, \tag{3.5}$$

where $\eta, r > 0$ and $\eta < \min[1 - \lambda_{(\overline{S},p)}, 1 - \overline{L}_{i_0}]$.

By Lemma 2.3 there exists a probability vector $(q_1, \ldots, q_N)$ such that $q_i : X \to [0,1]$ is Lipschitzian for $i \in \{1, \ldots, N\}$ and

$$q_i(x) > 0, \tag{3.6}$$

$$\max_{x \in X} |p_i(x) - q_i(x)| < \frac{\varepsilon}{2N} \quad \text{for } i = 1, \ldots, N. \tag{3.7}$$

Consider now the learning system $(T_1, \ldots, T_N; q_1, \ldots, q_N)$, where $T_i = \overline{S}_i$ for $i \in \{1, \ldots, N\}$, $i \neq i_0$ and $T_{i_0} = \overline{T}$. It follows immediately from (3.1), (3.4) and (3.7) that

$$d((S,p),(T,q)) < \varepsilon.$$

From (3.3), (3.6) and the fact that $q_i$ is Lipschitzian for $i \in \{1, \ldots, N\}$, according to Remark 2.1, $(T,q)$ is asymptotically stable. On the other hand, from (3.3) and (3.5) it follows that there exists an integer $n$ such that

$$T_{i_0}^n(X) = \{x_0\}.$$

Let $P$ be the Markov operator corresponding to $(T,q)$ and let $\mu_* \in \mathcal{M}_1$ be its unique invariant measure. We have

$$\begin{aligned}
\mu_*(\{x_0\}) &= (P^n \mu_*)(\{x_0\}) \\
&= \sum_{i_1, \ldots, i_n = 1}^{N} \int_X q_{i_1}(x) \ldots (q_{i_n} \circ T_{i_{n-1}} \circ \ldots \circ T_{i_1})(x) \\
&\qquad \times 1_{\{x_0\}}(T_{i_n} \circ \ldots \circ T_{i_1})(x)\, \mu_*(dx) \\
&\geq (\min_{x \in X} q_{i_0}(x))^n \mu_*(X) = (\min_{x \in X} q_{i_0}(x))^n.
\end{aligned}$$

By Lemma 2.1, $\mu_*$ is singular. Consequently, $(T,q) \in \mathcal{F}_0$. ∎

Remark 3.1. Suppose that $(S,p) \in \mathcal{F}_0$. Then there exists $l \geq 1$ such that the adjoint $U_{(S,P)}$ of $P_{(S,p)}$ satisfies the condition

$$U_{(S,p)} f \in F_l \quad \text{for } f \in F_l. \tag{3.8}$$

Proof. Let $r = \max_{1 \leq i \leq N} \operatorname{Lip} p_i$. A simple calculation shows that (3.8) holds for

$$l = \max\left[\frac{r}{1 - \lambda_{(S,p)}}, 1\right]. \ ∎$$

Lemma 3.2. *Let $(S,p) \in \mathcal{F}_0$. Then for all $\varepsilon > 0$, $l \geq 1$ and $n \in \mathbb{N}$ there exists $\delta > 0$ such that for each $(T,q) \in \mathcal{F}$,*

$$d((S,p),(T,q)) < \delta \Longrightarrow \sup_{f \in F_l,\, x \in X} |U^n_{(S,p)} f(x) - U^n_{(T,q)} f(x)| < \varepsilon.$$

Proof. Fix $\varepsilon > 0$, $n \in \mathbb{N}$ and $(S,p) \in \mathcal{F}_0$. From Remark 3.1 we have $U_{(S,p)} f \in F_t$ for $f \in F_t$, for some $t \geq 1$. Obviously $F_1 \subset F_t$. Let $f \in F_1$ and

$x \in X$. Then

$$
\begin{aligned}
(3.9)\qquad & |U_{(S,p)}f(x) - U_{(T,q)}f(x)| \\
&= \Big|\sum_{i=1}^{N} p_i(x)f(S_i(x)) - \sum_{i=1}^{N} q_i(x)f(T_i(x))\Big| \\
&\le \sum_{i=1}^{N} |p_i(x) - q_i(x)| + \sum_{i=1}^{N} q_i(x)|f(S_i(x)) - f(T_i(x))| \\
&\le \sum_{i=1}^{N} |p_i(x) - q_i(x)| + \sum_{i=1}^{N} \|S_i(x) - T_i(x)\| \le d((S,p),(T,q)).
\end{aligned}
$$

For $m > 1$ we have

$$
\begin{aligned}
|U^m_{(S,p)}f(x) - U^m_{(T,q)}f(x)| &\le |U_{(T,q)}(U^{m-1}_{(S,p)}f)(x) - U_{(T,q)}(U^{m-1}_{(T,q)}f)(x)| \\
&\quad + |U_{(T,q)}(U^{m-1}_{(S,p)}f)(x) - U_{(S,p)}(U^{m-1}_{(S,p)}f)(x)| \\
&\le \sup_{f\in F_1,\, y\in X} |U^{m-1}_{(S,p)}f(y) - U^{m-1}_{(T,q)}f(y)| \\
&\quad + |U_{(T,q)}(U^{m-1}_{(S,p)}f)(x) - U_{(S,p)}(U^{m-1}_{(S,p)}f)(x)|.
\end{aligned}
$$

Since $U^{m-1}_{(S,p)}f \in F_t$, we have $U^{m-1}_{(S,p)}f/t \in F_1$ and the last inequality can be written as

$$
\begin{aligned}
|U^m_{(S,p)}f(x) - U^m_{(T,q)}f(x)| &\le \sup_{f\in F_1,\, x\in X}\{|U^{m-1}_{(S,p)}f(x) - U^{m-1}_{(T,q)}f(x)|\} \\
&\quad + t \sup_{f\in F_1,\, x\in X}\{|U_{(S,p)}f(x) - U_{(T,q)}f(x)|\}.
\end{aligned}
$$

This and the inequality (3.9) yield

$$
\begin{aligned}
(3.10)\qquad & \sup_{f\in F_1,\, x\in X}\{|U^m_{(S,p)}f(x) - U^m_{(T,q)}f(x)|\} \\
&\le \sup_{f\in F_1,\, x\in X}\{|U^{m-1}_{(S,p)}f(x) - U^{m-1}_{(T,q)}f(x)|\} + td((S,p),(T,q))
\end{aligned}
$$

and consequently

$$
\sup_{f\in F_1,\, x\in X}\{|U^n_{(S,p)}f(x) - U^n_{(T,q)}f(x)|\} \le tnd((S,p),(T,q)).
$$

If $f \in F_l$ for $l \ge 1$, then $f/l \in F_1$ and we get

$$
\sup_{f\in F_l,\, x\in X}\{|U^n_{(S,p)}f(x) - U^n_{(T,q)}f(x)|\} \le tnld((S,p),(T,q)).
$$

Set $\delta = \varepsilon(tnl)^{-1}$ to complete the proof. ∎

Remark 3.2. For all $\varepsilon > 0$, $l \geq 1$ and $f \in C(X)$ there exists $\delta > 0$ such that for any $\mu_1, \mu_2 \in \mathcal{M}_1$,

$$\|\mu_1 - \mu_2\|_l < \delta \Longrightarrow |\langle f, \mu_1\rangle - \langle f, \mu_2\rangle| < \varepsilon.$$

Proof. Let $\varepsilon > 0$ and $f \in C(X)$. By Stone's Theorem there exists a Lipschitz function $\overline{f} : X \to \mathbb{R}$ satisfying

$$\max_{x\in X} |f(x) - \overline{f}(x)| < \varepsilon/3.$$

Without any loss of generality, we can assume that its Lipschitz constant $\overline{L}$ satisfies $\overline{L} \geq 1$. Let $\delta = \varepsilon(3\overline{L})^{-1}$. Since $l \geq 1$, we have $\overline{f}/\overline{L} \in F_l$. Therefore,

$$|\langle \overline{f}/\overline{L}, \mu_1\rangle - \langle \overline{f}/\overline{L}, \mu_2\rangle| \leq \|\mu_1 - \mu_2\|_l < \delta$$

and we obtain

$$|\langle \overline{f}, \mu_1\rangle - \langle \overline{f}, \mu_2\rangle| < \varepsilon/3.$$

Consequently, we get

$$\begin{aligned}|\langle f, \mu_1\rangle - \langle f, \mu_2\rangle| &\leq |\langle \overline{f}, \mu_1\rangle - \langle \overline{f}, \mu_2\rangle| + |\langle f - \overline{f}, \mu_1\rangle| \\ &\quad + |\langle f - \overline{f}, \mu_2\rangle| \leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \ \blacksquare\end{aligned}$$

We are now in a position to prove the main result of our paper.

Theorem 3.1. *The set $\overline{\mathcal{F}}_0$ of all $(S,p) \in \mathcal{F}$ which are asymptotically stable and have a singular stationary distribution is residual in $\mathcal{F}$.*

Proof. Let $(f_i)_{i\geq 1}$ be a sequence dense in $C(X)$. Fix $n\in\mathbb{N}$ and $(S,p)\in \mathcal{F}_0$. Let $P_{(S,p)}$ be the Markov operator corresponding to $(S,p)$ and $\mu_{(S,p)}$ its stationary distribution. Since $\mu_{(S,p)}$ is singular, we can consider a compact set $F_{(S,p),n} \subset X$ such that

$$\mu_{(S,p)}(F_{(S,p),n}) \geq 1 - \frac{1}{2n} \quad \text{and} \quad m(F_{(S,p),n}) = 0. \tag{3.11}$$

Further, due to regularity of the Lebesgue measure we can find a positive number $r_{(S,p),n}$ such that

$$m(B(F_{(S,p),n}, r_{(S,p),n})) \leq \frac{m(X)}{2n}, \tag{3.12}$$

where $B(F_{(S,p),n}, r_{(S,p),n})$ is the $r_{(S,p),n}$-neighbourhood of $F_{(S,p),n}$. Define $A_{(S,p),n} := X \setminus B(F_{(S,p),n}, r_{(S,p),n})$. From (3.12) the set $A_{(S,p),n}$ is nonempty and we can consider the Tietze function $h_{(S,p),n} : X \to \mathbb{R}_+$ given by the formula

$$h_{(S,p),n} = \frac{\|x, A_{(S,p),n}\|}{\|x, A_{(S,p),n}\| + \|x, F_{(S,p),n}\|},$$

where $\|x, A\|$ stands for the distance of the point $x$ from the set $A$ for $A \subset X$. It is easy to check that $h_{(S,p),n}(x) = 0$ for $x \notin B(F_{(S,p),n}, r_{(S,p),n})$, and $h_{(S,p),n}(x) = 1$ for $x \in F_{(S,p),n}$. For every $(S,p) \in \mathcal{F}_0$ and $n \in \mathbb{N}$ we will define the values $l \geq 1$, $k \in \mathbb{N}$, $\varepsilon > 0$ and $\delta_{(S,p),n} > 0$. Namely, by Remark 3.1 there exists $l \geq 1$ such that

$$\|P_{(S,p)}\mu_1 - P_{(S,p)}\mu_2\|_l \leq \|\mu_1 - \mu_2\|_l \quad \text{for } \mu_1, \mu_2 \in \mathcal{M}_1. \tag{3.13}$$

Using Remark 3.2 we can find $\varepsilon > 0$ such that for all $\mu_1, \mu_2 \in \mathcal{M}_1$,

$$\|\mu_1 - \mu_2\|_l < \varepsilon \Longrightarrow |\langle f, \mu_1\rangle - \langle f, \mu_2\rangle| < \frac{1}{2n} \tag{3.14}$$

for $f \in \{f_1, \ldots, f_n, h_{(S,p),n}\}$. It follows from Proposition 2.1 and Remark 2.1 that there is an integer $k$ such that

$$\|P^k_{(S,p)}\mu_1 - P^k_{(S,p)}\mu_2\|_l \leq \varepsilon/3 \quad \text{for } \mu_1, \mu_2 \in \mathcal{M}_1. \tag{3.15}$$

It follows from Lemma 3.2 that there is $\delta_{(S,p),n} > 0$ such that for all $(T,q) \in \mathcal{F}$,

$$\begin{aligned} d((S,p),(T,q)) < \delta_{(S,p),n} \\ \Longrightarrow \sup_{f \in F_l,\, x \in X} |U^k_{(S,p)}f(x) - U^k_{(T,q)}f(x)| < \varepsilon/3. \end{aligned} \tag{3.16}$$

Define

$$\overline{\mathcal{F}} = \bigcap_{n=1}^{\infty} \bigcup_{(S,p) \in \mathcal{F}_0} B_{\mathcal{F}}((S,p), \delta_{(S,p),n}),$$

where $B_{\mathcal{F}}((S,p), \delta_{(S,p),n})$ is the open ball in $(\mathcal{F}, d)$ with centre $(S,p)$ and radius $\delta_{(S,p),n}$. Obviously $\overline{\mathcal{F}}$ as an intersection of countably many open dense sets is residual. We are going to show that $\overline{\mathcal{F}} \subset \overline{\mathcal{F}}_0$.

Fix $(T,q) \in \overline{\mathcal{F}}$. Let $P_{(T,q)}$ denote the Markov operator corresponding to $(T,q)$. Since $X$ is compact, the operator $P_{(T,q)}$ has a stationary distribution $\mu_{(T,q)}$ [7], and to prove the asymptotic stability of $P_{(T,q)}$, is enough to check the weak convergence of the sequence $(P^n_{(T,q)}\mu)_{n\geq 1}$ to $\mu_{(T,q)}$, i.e.

$$\lim_{n\to\infty} \langle f, P^n_{(T,q)}\mu\rangle = \langle f, \mu_{(T,q)}\rangle \quad \text{for } f \in C(X) \text{ and } \mu \in \mathcal{M}_1. \tag{3.17}$$

Assume for contradiction that the formula does not hold. Then there exist $f_{i_0} \in C(X)$, $\mu \in \mathcal{M}_1$, and an increasing sequence $(n_m)_{m\geq 1}$ of integers such that

$$|\langle f_{i_0}, P^{n_m}_{(T,q)}\mu\rangle - \langle f_{i_0}, \mu_{(T,q)}\rangle| \geq \eta \tag{3.18}$$

for some $\eta > 0$ and all $m \geq 1$.

Choose $n_0 \in \mathbb{N}$ so large that $1/n_0 \leq \eta$ and $n_0 > i_0$. Since $(T,q) \in \overline{\mathcal{F}}$ it follows that $(T,q) \in B_{\mathcal{F}}((S,p), \delta_{(S,p),n_0})$ for some $(S,p) \in \mathcal{F}_0$. Assume that $l \geq 1, k \in \mathbb{N}$ and $\varepsilon > 0$ are such that conditions (3.13)–(3.15) hold for $P_{(S,p)}$ and $n = n_0 \in \mathbb{N}$.

Fix $n \geq k$ and set $\overline{\mu} := P^{n-k}_{(T,q)}\mu$. By the definition of $\delta_{(S,p),n_0}$ we get

$$\begin{aligned}\|P^n_{(T,q)}\mu - \mu_{(T,q)}\|_l = \|P^k_{(T,q)}\overline{\mu} - P^k_{(T,q)}\mu_{(T,q)}\|_l &\leq \|P^k_{(S,p)}(\overline{\mu} - \mu_{(T,q)})\|_l \\ &+ \|(P^k_{(S,p)} - P^k_{(T,q)})\overline{\mu}\|_l + \|(P^k_{(S,p)} - P^k_{(T,q)})\mu_{(T,q)}\|_l \\ &\leq \varepsilon/3 + 2 \sup_{f\in F_l,\, x\in X} |U^k_{(S,p)}f(x) - U^k_{(T,q)}f(x)| \leq \varepsilon.\end{aligned}$$

Hence for $n \geq k$ we have

$$|\langle f, P^n_{(T,q)}\mu\rangle - \langle f, \mu_{(T,q)}\rangle| < \frac{1}{2n_0} \quad \text{for } f \in \{f_1, \ldots, f_{n_0}, h_{(S,p),n_0}\}.$$

Since $n_0 > i_0$ it follows that for $n_m \geq k$ we have

$$|\langle f_{i_0}, P^{n_m}_{(T,q)}\mu\rangle - \langle f_{i_0}, \mu_{(T,q)}\rangle| < \frac{1}{2n_0} < \eta.$$

This contradicts condition (3.18).

We only need to show that $\mu_{(T,q)} \in \mathcal{M}_1$ is singular. Let $((S,p)_n)_{n\geq 1}$ be a sequence of learning systems of $\mathcal{F}_0$ such that

$$(T,q) \in B_{\mathcal{F}}((S,p)_n, \delta_{(S,p)_n,n}) \quad \text{for } n \in \mathbb{N}.$$

Denote by $\mu_{(S,p)_n}$ the stationary distribution of the operator $P_{(S,p)_n}$. Assume that $l_n \geq 1$, $k_n \in \mathbb{N}$ and $\varepsilon_n > 0$ are such that (3.13)–(3.15) hold for $P_{(S,p)_n}$ and $n \in \mathbb{N}$. Hence

$$\begin{aligned}\|\mu_{(T,q)} - \mu_{(S,p)_n}\|_{l_n} &= \|P^{k_n}_{(T,q)}\mu_{(T,q)} - P^{k_n}_{(S,p)_n}\mu_{(S,p)_n}\|_{l_n} \\ &\leq \|P^{k_n}_{(S,p)_n}\mu_{(T,q)} - P^{k_n}_{(S,p)_n}\mu_{(S,p)_n}\|_{l_n} + \|(P^{k_n}_{(T,q)} - P^{k_n}_{(S,p)_n})\mu_{(T,q)}\|_{l_n}.\end{aligned}$$

By the definitions of $l_n$ and $k_n$, and the above estimate,

$$\|\mu_{(T,q)} - \mu_{(S,p)_n}\|_{l_n} < \frac{2}{3}\varepsilon_n \quad \text{for } n \in \mathbb{N}, \tag{3.20}$$

where $\varepsilon_n > 0$ is such that for all $\mu_1, \mu_2 \in \mathcal{M}_1$ the implication

$$\|\mu_1 - \mu_2\|_l < \varepsilon_n \Longrightarrow |\langle f, \mu_1\rangle - \langle f, \mu_2\rangle| < \frac{1}{2n}$$

holds for $f \in \{f_1, \ldots, f_n, h_{(S,p)_n,n}\}$. It follows that for every $n \in \mathbb{N}$,

$$|\langle h_{(S,p)_n,n}, \mu_{(T,q)}\rangle - \langle h_{(S,p)_n,n}, \mu_{(S,p)_n}\rangle| < \frac{1}{2n}$$

and by the definition of $h_{(S,p)_n,n}$ we get

$$\begin{aligned}\mu_{(T,q)}(B(F_{(S,p)_n,n}, r_{(S,p)_n,n})) &> \mu_{(S,p)_n}(F_{(S,p)_n,n}) - \frac{1}{2n} \\ &\geq 1 - \frac{1}{2n} - \frac{1}{2n} = 1 - \frac{1}{n}.\end{aligned}$$

Since (3.12) holds, by Lemma 2.1 we conclude that $\mu_{(T,q)}$ is singular. ■

**References**

[1] W. Bartoszek, *Norm residuality of ergodic operators*, Bull. Polish Acad. Sci. Math. 29 (1981), 165–167.
[2] R. M. Dudley, *Probabilities and Metrics*, Aarhus Universitet, 1976.
[3] R. Fortet et B. Mourier, *Convergence de la répartition empirique vers la répartition théorétique*, Ann. Sci. École Norm. Sup. 70 267–285 (1953).
[4] A. Iwanik, *Approximation theorem for stochastic operators*, Indiana Univ. Math. J. 29 (1980), 415–425.
[5] A. Iwanik and R. Rębowski, *Structure of mixing and category of complete mixing for stochastic operators*, Ann. Polon. Math. 56 (1992), 233–242.
[6] A. Lasota and J. Myjak, *Generic properties of fractal measures*, Bull. Polish Acad. Sci. Math. 42 (1994), 283–296.
[7] A. Lasota and J. A. Yorke, *Lower bound technique for Markov operators and iterated function systems*, Random Comput. Dynam. 2 (1994), 41–77.
[8] T. Szarek, *Generic properties of continuous iterated function systems*, Bull. Polish Acad. Sci. Math. 47 (1997), 77–89.
[9] —, *Iterated function systems depending on a previous transformation*, Univ. Iagel. Acta Math. 33 (1996), 161–172.
[10] —, *Markov operators acting on Polish spaces*, Ann. Polon. Math. 67 (1997), 247–257.

Institute of Mathematics
Polish Academy of Sciences
Staromiejska 8/6
40-013 Katowice, Poland
E-mail: szarek@gate.math.us.edu.pl

*Reçu par la Rédaction le 15.6.1998*
*Révisé le 24.3.1999 et 12.8.1999*