Proc. Convergence in Ergodic Theory and Probability (Columbus 1993),
Ohio State Univ. Math. Res. Inst. Publ. Vol. 5, Walter de Gruyter, Berlin, 1996, (305-332).
Preprint Ser. No. 14, 1993, *Math. Inst. Aarhus.*

The Uniform Ergodic Theorem for Dynamical Systems

GORAN PESKIR* *and* MICHEL WEBER

Necessary and sufficient conditions are given for the uniform convergence over an arbitrary index set in von Neumann's mean and Birkhoff's pointwise ergodic theorem. Three different types of conditions already known from probability theory are investigated. Firstly it is shown that the property of being eventually totally bounded in the mean is necessary and sufficient. This condition involves as a particular case Blum-DeHardt's theorem which offers the best known sufficient condition for the uniform law of large numbers in the independent case. Secondly it is shown that eventual tightness is necessary and sufficient. In this way a link with a weak convergence is obtained. Finally it is shown that the existence of some particular totally bounded pseudo-metrics is necessary and sufficient. The conditions derived are of Lipschitz type, while the method of proof relies upon a result of independent interest, called the uniform ergodic lemma. This result considerably extends Hopf-Yosida-Kakutani's maximal ergodic lemma to a form more suitable for examinations of the uniform convergence under consideration. From this lemma an inequality is also derived which extends the classical (weak) maximal ergodic inequality to the uniform case. In addition, a uniform approximation by means of a dense family of maps satisfying the uniform ergodic theorem in a trivial way is investigated, and a particular result of this type is established. This approach is in the spirit of the classical Hilbert space method for the mean ergodic theorem of von Neumann, and therefore from the ergodic theory point of view it could be seen as the natural one. After this, a simple characterization is obtained for the uniform convergence of moving averages. Finally, a counter-example is constructed for a symmetrization inequality in the stationary ergodic case. This inequality is known to be of vital importance to support the Vapnik-Chervonenkis random entropy approach in the independent case. Further developments in this direction are indicated.

1. Introduction

What is nowadays called the *uniform ergodic theorem* goes back to 1941 when Yosida and Kakutani published the study [52]. In this study they obtained conditions[1] on a bounded linear operator T in a Banach space B which are sufficient to provide that the averages $n^{-1}\sum_{j=0}^{n-1} T^j$ converge in the uniform operator topology to an operator P in B :

$$\Big\| \, \frac{1}{n} \sum_{j=0}^{n-1} T^j - P \, \Big\| \longrightarrow 0 \tag{1.1}$$

as $n \rightarrow \infty$. In this context it should be recalled that the operator T satisfying (1.1) is called

* Research partially supported by Danish Natural Science Research Council and partially by Danish Research Academy.
AMS 1980 *subject classifications*. Primary 28D05, 47A35, 60G10. Secondary 28A20, 60B12, 60F15, 60F25.
Key words and phrases: Uniform ergodic theorem, the uniform ergodic lemma (inequality), the prime problem, the dual problem, eventually totally bounded in the mean, eventually tight, totally bounded, Blum-DeHardt's theorem, the VC random entropy numbers, positive contraction, dynamical system, stationary, ergodic, measure-preserving, perfect, mixing, moving average, Banach limit. © goran@imf.au.dk

[1] If T is power bounded and quasi-compact, then (1.1) holds true (see [25] p.86-92).

uniformly ergodic (see [25] p.86). Their work generalizes previously established results of Fréchet and Visser, and in particular of Krylov and Bogolioubov whose concept of *compact operator* is shown to be of vital importance in this direction. Here we find it convenient to recall the words of Yosida and Kakutani which describe how the uniform ergodic theorem has been discovered. They wrote (see [52] p.195): "… *This condition*[1] *was investigated by J. L. Doob*[2] *without being noticed that the linear operation* T *becomes weakly completely continuous*[3] *under this condition. We have once*[4] *treated this case of the condition of J. L. Doob as an application of the mean ergodic theorem. But, on looking precisely into the detail of the fact, we have found that, in this case, the linear operation* T *satisfies even the (in some sense stronger) condition*[5] *of N. Kryloff-N. Bogolioùboff, and that the uniform ergodic theorem is true in this case. (Indeed, the condition of N. Kryloff-N. Bogolioùboff follows from the condition*[6] *of W. Doeblin, which is weaker than that of J. L. Doob.)* …"

In the rest of the Yosida-Kakutani study the result has been well applied to some problems in the theory of Markov processes. A number of studies have followed. Among those we point out [1], [2], [4], [5], [14], [19], [23][7], [24], [30][8], [31], [32][9], [33], [34], [43], as well as the presentation of the Yosida-Kakutani result in [13] (p.708-717) and [25] (p.86-94) where further references can be found. Of the particular Banach spaces that were under consideration, it seems that the case of the Banach spaces $C(S)$ and $L_1(S,\Sigma,\mu)$ in the notation from [13] has been studied in more detail (see [1], [2], [4], [13], [23], [34]). Less seems to be known in this context about some of the other spaces.

On the other hand in probability theory nowadays, we have the fundamental *Glivenko-Cantelli theorem* which goes back to 1933 when papers [6] and [18] were published. A large number of studies were followed. As general references we point out [11], [15], [41] and [42] where additional references can be found. In the recent years various extensions of the classical Glivenko-Cantelli theorem have appeared. They could be commonly called *uniform laws of large numbers*. The main problem under considerations in this context, roughly speaking, may be stated as follows. Given a sequence of independent and identically distributed random variables $\{\, \xi_j \mid j \geq 1 \,\}$ with values in a space S, and a map f from $S \times \Theta$ into $\mathbf{R}$, determine conditions under which the uniform convergence is valid:

$$(1.2) \qquad \sup_{\theta \in \Theta} \Big| \, \frac{1}{n} \sum_{j=1}^{n} f(\xi_j,\theta) - M(\theta) \, \Big| \longrightarrow 0$$

as $n \to \infty$, where $M(\theta)$ is the average of $f(\xi_1,\theta)$ for $\theta \in \Theta$, and the convergence in (1.2) is one of the standard probabilistic ones. It seems apparent at the present time that two distinct

[1] A uniform integrability condition which characterizes weak compactness of the (integral) operator under consideration.

[2] It has been done in [10].

[3] It means *weakly compact* in today's language.

[4] It was in [50].

[5] It is *quasi-compactness* (in today's language) which was introduced in [26] and [27].

[6] There exist $d \geq 1$ and ε, $\delta > 0$ such that $\mu(A) < \delta$ implies $P^{(d)}(t,A) \leq 1-\varepsilon$ for all t. (Here $P(t,A)$ is a transition probability.) For more details see Doeblin's paper [9].

[7] If T is a conservative and ergodic positive contraction of L^1, then it is uniformly ergodic if and only if it is quasi-compact.

[8] If $\|T^n\|/n \to 0$, then T is uniformly ergodic if and only if $(I-T)(B)$ is closed.

[9] If T is a positive operator on a Banach lattice with $\|T^n\|/n \to 0$, then T is quasi-compact if (and only if) the averages of its iterates converge uniformly to a finite-dimensional projection.

approaches towards solution for this problem have emerged. The first one offers Lipschitz type conditions and relies upon *Blum-DeHardt's law of large numbers* that is established in [3] and [7]. This law offers presently the best known sufficient condition for (1.2) to be valid. Subsequently, a necessary and sufficient condition for (1.2), called *eventually total boundedness in the mean*, is obtained in [20], and although not explicitly recorded there, this condition involves Blum-DeHardt's law of large numbers as a particular case. That result has been recently shown to be valid in the stationary ergodic case as well (see [37]). This fact clearly indicates that further extensions of this approach into ergodic theory might be possible. (For a related work we refer the reader to [21], and in particular to [36] where it is shown that *Hardy's regular convergence* of certain means is necessary and sufficient for the most useful consequence of (1.2) in statistics.) The second approach has been developed by *Vapnik and Chervonenkis* through the fundamental studies [45] and [46]. It offers combinatorial conditions in terms of random entropy numbers. In this direction we point out papers [17] and [53] as well as the presentation in [29] (p.394-420). Let us also mention that somewhat different necessary and sufficient conditions for (1.2) are obtained in [44].

It is the purpose of the present paper to point out the closeness of problems (1.1) and (1.2), and to obtain their unification through a single approach. Let us for this reason look more closely at (1.1) and (1.2) itself. First it should be noted that (1.1) does not provide any application to (1.2). Moreover, if $B = L^1(\mu)$ and T is the composition with a measure-preserving transformation, then T satisfies (1.1) if and only if T is periodic with bounded period. Similarly, it seems that having any particular version of (1.2), the best we can deduce in the direction of (1.1) is the mean ergodic theorem, that is the pointwise convergence on B of the averages from (1.1). This should follow from uniform integrability of averages along the lines of [11] (p.42-43), at least for T being the composition with a stationary ergodic shift in a countable product of copies of a separable Banach space, and with Θ being a norming subset of the unit ball in the dual space (see also [11] p.3). Therefore it seems to be of interest to reconsider the uniform ergodic theorem (1.1) as follows. We suppose that T and P are maps on some space B (which is (are) to be specified from the character of the objects being involved).

Prime problem. *Determine conditions which are* (*necessary and*) *sufficient for the convergence to be valid:*

$$\frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\theta) - P(f_\theta) \longrightarrow 0 \ \text{ uniformly in } \ \theta \in \Theta$$

as $n \to \infty$ *, where* f_θ *are with values in* B *for* $\theta \in \Theta$ *.*

Dual Problem. *Determine conditions which are* (*necessary and*) *sufficient for the convergence to be valid:*

$$\frac{1}{n}\sum_{j=0}^{n-1} f_\theta(T^j) - f_\theta(P) \longrightarrow 0 \ \text{ uniformly in } \ \theta \in \Theta$$

as $n \to \infty$ *, where* f_θ *are with arguments in* B *for* $\theta \in \Theta$ *.*

These two problems seem to be of different nature, but each is interesting from its own standpoint. For instance, taking for Θ in the prime problem the unit ball in a Banach space B and norm convergence, we obtain the uniform ergodic theorem (1.1). Similarly, taking for Θ in the dual problem the unit ball in its dual space B^* and weak*-convergence, we obtain the mean ergodic theorem for T in B . It should be noted here that in the convergence statements in the prime problem and in the dual problem we did not specify which convergence takes the place. We take this liberty to clarify the main ideas of unification, and at the same time to leave various possibilities open. For instance, if B is a function space, then the expression on the left-hand side in the prime problem could define a function with values in a normed space. Thus the question as to which convergence takes the place in the prime problem depends on the norm we consider, due to the fact that in this case instead of the function norm we can consider a norm in the range space of the function space as well. The same remark might be directed to the dual problem. Moreover, replacing the unit ball of a Banach space B in the prime problem by a smaller family of vectors (that may be of some interest) allows the uniform ergodic theorem (1.1) to become weaker and at the same time more easily established (see Example 4.6). In this way the anomaly with measure-preserving transformations stated above could be avoided. Finally, it is obvious that these two problems become mutually equivalent in the case where the operator T is induced by a measure-preserving transformation (at least when this transformation is invertible), and therefore they involve (1.2) as a particular case. This is precisely the case which is mainly considered and investigated in the paper.

We shall say just a few words about the organization of the paper, and rather concentrate to the results in a straightforward way. In the next section we prove a result which is of a considerable interest in itself. We call it the *uniform ergodic lemma* (for positive contractions), and to the best of our knowledge this sort of result has not been studied or recorded previously. Using this result in the third section we offer solutions for the prime problem and dual problem in the case where the operator T is induced by a measure-preserving transformation. In this context it should be recalled that the pointwise ergodic theorem of Birkhoff is proved by using the maximal ergodic lemma, while now the uniform ergodic theorem for dynamical systems is proved by using the uniform ergodic lemma. We think that this fact is by itself of theoretical interest. In the last section we present another approach towards the uniform ergodic theorem for dynamical systems (which is in the spirit of the classical Hilbert space method for the mean ergodic theorem of von Neumann) and various examples.

2. The uniform ergodic lemma

The proof of the forthcoming uniform ergodic theorem in the next section relies upon a fact of independent interest that is presented in Theorem 2.1 below. This result is instructive to be compared with Hopf-Yosida-Kakutani's maximal ergodic lemma (see [25] p.8). In Corollary 2.2 below we present its consequence which should be compared with the maximal ergodic inequality (see [16] p. 24,29). We begin by introducing the notation and recalling some facts on the non-measurable calculus needed in the sequel.

Given a linear operator T in $L^p(\mu)$ and $f \in L^p(\mu)$ for some $1 \leq p < \infty$, we denote:

(2.1) $$S_n(f) = \sum_{j=0}^{n-1} T^j f$$

(2.2) $$M_n(f) = \max_{1 \leq j \leq n} S_j(f)$$

(2.3) $$R_n(f) = \max_{1 \leq j \leq n} S_j(f)/j$$

for $n \geq 1$. We shall restrict our attention to the case where the underlying measure space $(X, \mathcal{A}, \mu)$

is σ-finite, but we remark that further extensions are possible. The symbols μ^* and μ_* denote the*outer μ-measure* and the *inner μ-measure* respectively. The *upper μ-integral* of an arbitrary function f from X into $\bar{\mathbf{R}}$ is defined as follows $\int^* f\,d\mu = \inf\,\{\,\int g\,d\mu \mid g \in L^1(\mu)\,, f \le g\,\}$, with the convention $\inf\emptyset = +\infty$. The *lower μ-integral* of an arbitrary function f from X into $\bar{\mathbf{R}}$ is defined as follows $\int_* f\,d\mu = \sup\,\{\,\int g\,d\mu \mid g \in L^1(\mu)\,,\, g \le f\,\}$, with the convention $\sup\emptyset = -\infty$. We denote by f^* the *upper μ-envelope* of f . This means that f^* is an $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $f \le f^*$, and if g is an another $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $f \le g$ μ-a.e., then $f^* \le g$ μ-a.e. We denote by f_* the *lower μ-envelope* of f . This means that f_* is an $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $f_* \le f$, and if g is an another $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $g \le f$ μ-a.e., then $g \le f_*$ μ-a.e. It should be noted that such envelopes exist under the assumption that μ is σ-finite. It is well-known that we have $\int^* f\,d\mu = \int f^*\,d\mu$, whenever the integral on the right-hand side exists in $\bar{\mathbf{R}}$, and $\int^* f\,d\mu = +\infty$ otherwise. Similarly we have $\int_* f\,d\mu = \int f_*\,d\mu$, whenever the integral on the right-hand side exists in $\bar{\mathbf{R}}$, and $\int_* f\,d\mu = -\infty$ otherwise. For more information in this direction we refer to [35]. To conclude the preliminary part of the section, we clarify that $\int_A^* f\,d\mu$ stands for $\int^* f\cdot 1_A\,d\mu$, whenever $A \subset X$ is an arbitrary set and $f : X \to \bar{\mathbf{R}}$ is an arbitrary function.

Theorem 2.1 (**The uniform ergodic lemma**)
Let T be a positive contraction in $L^1(\mu)$, and let $\{\,f_\theta \mid \theta \in \Theta\}$ be a family of functions from $L^1(\mu)$. Let us denote $A_n = \{\,\sup_{\theta\in\Theta} M_n(f_\theta) > 0\,\}$ and $B_n = \{\,(\sup_{\theta\in\Theta} M_n(f_\theta))^* > 0\,\}$ for all $n \ge 1$. Then we have:

$$\int_{A_n}^* \sup_{\theta\in\Theta} f_\theta\ d\mu \ \ge\ 0 \tag{2.4}$$

$$\int_{B_n}^* \sup_{\theta\in\Theta} f_\theta\ d\mu \ \ge\ 0 \tag{2.5}$$

for all $n \ge 1$.

Proof. We shall first assume that $\sup_{\theta\in\Theta} f_\theta \le g$ for some $g \in L^1(\mu)$. In this case we have $(\sup_{\theta\in\Theta} f_\theta)^* \in L^1(\mu)$. Let $n \ge 1$ be given and fixed. By monotonicity of T we get:

$$S_j(f_\theta) \,=\, f_\theta + TS_{j-1}(f_\theta) \,\le\, f_\theta + T((M_n(f_\theta))^+) \tag{2.6}$$

for all $j = 2,\dots,n+1$, and all $\theta \in \Theta$. Moreover, since $T((M_n(f_\theta))^+) \ge 0$ for all $\theta \in \Theta$, we see that (2.6) is valid for $j = 1$ as well. Hence we find:

$$M_n(f_\theta) \,\le\, f_\theta + T((M_n(f_\theta))^+)$$

for all $\theta \in \Theta$. Taking supremum over all $\theta \in \Theta$ we obtain:

$$\sup_{\theta\in\Theta} M_n(f_\theta) \,\le\, \sup_{\theta\in\Theta} f_\theta + \sup_{\theta\in\Theta}\, T((M_n(f_\theta))^+) \ . \tag{2.7}$$

Since $f_\theta \le g$ for all $\theta \in \Theta$, then we have $\sup_{\theta\in\Theta} (M_n(f_\theta))^+ \le (M_n(g))^+$. Hence we see that

$(\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^* \in L^1(\mu)$. Therefore by (2.7) and monotonicity of T we get:

$$\sup_{\theta\in\Theta} M_n(f_\theta) \leq \sup_{\theta\in\Theta} f_\theta + T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ . \tag{2.8}$$

Multiplying both sides by 1_{A_n} we obtain:

$$\sup_{\theta\in\Theta} (M_n(f_\theta))^+ = (\sup_{\theta\in\Theta} M_n(f_\theta))^+ = \sup_{\theta\in\Theta} M_n(f_\theta) \cdot 1_{A_n} \leq$$

$$\sup_{\theta\in\Theta} f_\theta \cdot 1_{A_n} + T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \cdot 1_{A_n} \leq$$

$$\sup_{\theta\in\Theta} f_\theta \cdot 1_{A_n} + T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ .$$

Integrating both sides we get:

$$\int^* \sup_{\theta\in\Theta} (M_n(f_\theta))^+ \, d\mu \ \leq$$

$$\int_{A_n}^* \sup_{\theta\in\Theta} f_\theta \ d\mu \ + \int T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ d\mu \ .$$

Finally we may conclude:

$$\int^* \sup_{\theta\in\Theta} (M_n(f_\theta))^+ \, d\mu - \int T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ d\mu \leq \int_{A_n}^* \sup_{\theta\in\Theta} f_\theta \ d\mu$$

and the proof of (2.4) follows from contractibility of T . In addition from (2.8) we get:

$$(\sup_{\theta\in\Theta} M_n(f_\theta))^* \leq (\sup_{\theta\in\Theta} f_\theta)^* + T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ .$$

Multiplying both sides by 1_{B_n} we obtain:

$$(\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^* = ((\sup_{\theta\in\Theta} M_n(f_\theta))^+)^* =$$

$$((\sup_{\theta\in\Theta} M_n(f_\theta))^*)^+ = (\sup_{\theta\in\Theta} M_n(f_\theta))^* \cdot 1_{B_n} \leq$$

$$(\sup_{\theta\in\Theta} f_\theta)^* \cdot 1_{B_n} + T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \cdot 1_{B_n} \leq$$

$$(\sup_{\theta\in\Theta} f_\theta)^* \cdot 1_{B_n} + T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ .$$

Integrating both sides we get:

$$\int^* \sup_{\theta\in\Theta} (M_n(f_\theta))^+ \, d\mu \ \leq$$

$$\int_{B_n}^* \sup_{\theta\in\Theta} f_\theta \ d\mu \ + \int T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ d\mu \ .$$

Finally, as above, we may conclude:

$$\int^* \sup_{\theta\in\Theta} (M_n(f_\theta))^+ \, d\mu - \int T((\sup_{\theta\in\Theta} (M_n(f_\theta))^+)^*) \ d\mu \leq \int_{B_n}^* \sup_{\theta\in\Theta} f_\theta \ d\mu$$

and the proof of (2.5) follows from contractibility of T .

Next suppose that there is no $g \in L^1(\mu)$ satisfying $\sup_{\theta\in\Theta} f_\theta \le g$. In this case we have $\int^* \sup_{\theta\in\Theta} f_\theta \, d\mu = +\infty$. Let $n \ge 1$ be given and fixed, then by subadditivity we get:

$$\int^* \sup_{\theta\in\Theta} f_\theta \, d\mu \;\le\; \int_{C_n}^* \sup_{\theta\in\Theta} f_\theta \, d\mu \;+\; \int_{C_n^c}^* \sup_{\theta\in\Theta} f_\theta \, d\mu$$

with C_n equal to either A_n or B_n . However on the set C_n^c in both cases we evidently have $\sup_{\theta\in\Theta} f_\theta \le \sup_{\theta\in\Theta} M_n(f_\theta) \le 0$. Therefore $\int_{C_n}^* \sup_{\theta\in\Theta} f_\theta \, d\mu = +\infty$, and the proof of theorem is complete. □

Corollary 2.2 (**The uniform ergodic inequality**)

Under the hypotheses of Theorem 2.1 suppose moreover that μ *is finite, and that* $T(\mathbf{1}) = \mathbf{1}$ *. Let us denote* $A_{n,\alpha} = \{ \sup_{\theta\in\Theta} R_n(f_\theta) > \alpha \}$ *and* $B_{n,\alpha} = \{ (\sup_{\theta\in\Theta} R_n(f_\theta))^* > \alpha \}$ *for all* $n \ge 1$ *and all* $\alpha > 0$ *. Then we have:*

$$\int_{A_{n,\alpha}}^* \sup_{\theta\in\Theta} (f_\theta - \alpha) \, d\mu \ge 0 \tag{2.9}$$

$$\mu_*\{ \sup_{\theta\in\Theta} R_n(f_\theta) > \alpha \} \le \frac{1}{\alpha} \int_{A_{n,\alpha}}^* \sup_{\theta\in\Theta} f_\theta \, d\mu \tag{2.10}$$

$$\int_{B_{n,\alpha}}^* \sup_{\theta\in\Theta} (f_\theta - \alpha) \, d\mu \ge 0 \tag{2.11}$$

$$\mu^*\{ \sup_{\theta\in\Theta} R_n(f_\theta) > \alpha \} \le \frac{1}{\alpha} \int_{B_{n,\alpha}}^* \sup_{\theta\in\Theta} f_\theta \, d\mu \tag{2.12}$$

for all $n \ge 1$ *and all* $\alpha > 0$ *.*

Proof. Let $n \ge 1$ and $\alpha > 0$ be given and fixed. Consider $g_\theta = f_\theta - \alpha$ for $\theta \in \Theta$. Since $T(\mathbf{1}) = \mathbf{1}$, then it is easily verified that $A_{n,\alpha} = \{ \sup_{\theta\in\Theta} M_n(g_\theta) > 0 \}$ and therefore $B_{n,\alpha} = \{ (\sup_{\theta\in\Theta} M_n(g_\theta))^* > 0 \}$. Hence by subadditivity and Theorem 2.1 we obtain:

$$\int_{C_{n,\alpha}}^* \sup_{\theta\in\Theta} f_\theta \, d\mu \;+\; \int_{C_{n,\alpha}}^* (-\alpha) \, d\mu \;\ge\; \int_{C_{n,\alpha}}^* \sup_{\theta\in\Theta} g_\theta \, d\mu \;\ge\; 0$$

with $C_{n,\alpha}$ equal to either $A_{n,\alpha}$ or $B_{n,\alpha}$. Thus the proof follows straightforward from the facts that $\int_{A_{n,\alpha}}^* (-\alpha) \, d\mu = -\alpha \cdot \mu_*(A_{n,\alpha})$ and $\int_{B_{n,\alpha}}^* (-\alpha) \, d\mu = -\alpha \cdot \mu(B_{n,\alpha})$. □

Remark 2.3

Under the hypotheses of Theorem 2.1 and Corollary 2.2 respectively let us suppose that $(\Theta, \mathcal{B})$ is a *Blackwell space* [1] (see [22]). We remark that any *analytic space* [2], and thus any *polish space* [3], is a Blackwell space. Suppose moreover that the map $(x,\theta) \mapsto f_\theta(x)$ from $X \times \Theta$ into $\bar{\mathbf{R}}$ is measurable with respect to the product σ-algebra $\mathcal{A} \times \mathcal{B}$ and Borel σ-algebra $\mathcal{B}(\bar{\mathbf{R}})$. Then by the

[1] A measurable space $(\Theta, \mathcal{B})$ is called Blackwell, if $f(\Theta)$ is analytic (see below) for every measurable $f : \Theta \to \mathbf{R}$.

[2] An analytic space is a Hausdorff space which is a continuous image of a Polish space.

[3] A Polish space is a separable topological space that can be metrized by means of a complete metric.

projection theorem [4] (see [22]) we may conclude that the map $x \mapsto \sup_{\theta\in\Theta} f_\theta(x)$ is *universally $\mathcal{A}$-measurable* from X into $\bar{\mathbf{R}}$ (μ-measurable with respect to any measure μ on $\mathcal{A}$). In this way all upper μ-integrals (outer and inner μ-measures) in Theorem 2.1 and Corollary 2.2 become the ordinary ones (without stars) . Moreover, then both Theorem 2.1 and Corollary 2.2 extend to the case where supremum in the definitions of functions M and R is taken over all integers. We leave the formulation of these facts and remaining details to the reader.

3. The uniform ergodic theorem for dynamical systems

The main object under our consideration in this section is a given *ergodic* dynamical system $(X, \mathcal{B}, \mu, T)$. We clarify that $(X, \mathcal{B}, \mu)$ is supposed to be a probability space, and T is an ergodic measure-preserving transformation of X . In addition, we assume that a parameterized family $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ of measurable maps from X into $\mathbf{R}$ is given, such that the condition is satisfied:

$$\int^* \sup_{\theta\in\Theta} |f_\theta| \; d\mu < \infty \; . \tag{3.1}$$

In particular, we see that f_θ belongs to $L^1(\mu)$ for all $\theta \in \Theta$, and therefore from Birkhoff's pointwise ergodic theorem (see [25] p. 9) we have:

$$\frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) \longrightarrow M(\theta) \quad \mu\text{-a.s.} \tag{3.2}$$

as $n \to \infty$, for all $\theta \in \Theta$. The limit function M is called the *μ-mean function* of $\mathcal{F}$, and we have $M(\theta) = \int f_\theta \, d\mu$ for all $\theta \in \Theta$. Hence we see that M belongs to $B(\Theta)$ whenever (3.1) is satisfied, where $B(\Theta)$ denotes the set of all bounded real valued functions on Θ . It is the purpose of this section to present a solution for the *prime and dual problem* from Section 1 that now may be restated as follows. Determine conditions which are necessary and sufficient for the uniform convergence to be valid:

$$\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \longrightarrow 0 \tag{3.3}$$

as $n \to \infty$. More precisely, we shall consider three kinds of convergence notions in (3.3), namely $(a.s.)^*$-convergence, $(L^1)^*$-convergence, and (μ^*)-convergence. Given a sequence of arbitrary maps $\{ Z_n \mid n \geq 1 \}$ from X into $\mathbf{R}$, we recall that $Z_n \to 0$ $(a.s.)^*$ if $|Z_n|^* \to 0$ μ-a.s., that $Z_n \to 0$ $(L^1)^*$ if $|Z_n|^* \to 0$ in μ-mean, and that $Z_n \to 0$ (μ^*) if $|Z_n|^* \to 0$ in μ-measure. For more information in this direction we shall refer the reader to [36]. It should be noted that if all maps under consideration are measurable, then the convergence notions stated above coincide with the notion of μ-a.s. convergence, convergence in μ-mean, and convergence in μ-measure, respectively. This could be achieved in quite a general setting as it was already described in Remark 2.3. In order to keep generality and be able to handle measurability problems appearing in the sequel, we shall often assume that the transformation T is μ-perfect. This means

[4] Let $(X, \mathcal{A})$ be a measurable space, and let $(\Theta, \mathcal{B})$ be a Blackwell space. If π_X denotes the projection of $X \times \Theta$ onto X , then $\pi_X(C)$ is universally $\mathcal{A}$-measurable for every $C \in \mathcal{A} \times \mathcal{B}$. (This is a version of the projection theorem that suffices for our remark above. We remark that its further extensions and generalizations are available (see [22]).)

that for every $B \in \mathcal{B}$ there exists $C \in \mathcal{B}$ satisfying $C \subset T(B)$ and $\mu(B \setminus T^{-1}(C)) = 0$. We can require equivalently that $\mu^*(T^{-1}(C)) = 1$ whenever C is a subset of X satisfying $\mu^*(C) = 1$ (see [35]). Under this assumption we have (see [35]):

(3.4) $$(\psi \circ T)^* = \psi^* \circ T$$

(3.5) $$\int^* \psi \circ T \, d\mu = \int^* \psi \, d\mu$$

whenever $\psi : X \to \bar{\mathbf{R}}$ is an arbitrary map. Dynamical system $(X, \mathcal{B}, \mu, T)$ is said to be *perfect*, if T is μ-perfect. The best known sufficient condition for the μ-perfectness of T could be:

(3.6) $$T(B) \in \mathcal{B}_\mu \quad \text{for all} \quad B \in \mathcal{B}$$

where $\mathcal{B}_\mu$ denotes the μ-completion of $\mathcal{B}$. This condition is by the image theorem satisfied whenever $(X, \mathcal{B})$ is a countably separated Blackwell space (see [22]). Moreover, if $(X, \mathcal{B})$ equals to the countable product $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$ of copies of a measurable space $(S, \mathcal{A})$, then by the projection theorem we see that condition (3.6) is satisfied whenever $(S, \mathcal{A})$ is a Blackwell space (see [22]). We remind again that any analytic space is a Blackwell space, and that any analytic metric space, and thus any polish space, is a countably separated Blackwell space. To conclude the preliminary part of this section we remark that from the point of view of two distinct approaches towards solution for (3.3) described in Section 1, our approach below offers Lipschitz type conditions. For the random entropy combinatorial approach we shall refer the reader to [39]. Finally, it should be mentioned that our conditions on the dynamical system $(X, \mathcal{B}, \mu, T)$ and family $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ imposed above seem to be as optimal as possible from many points of view. However, we shall not present the details in this direction here, and for more detailed discussion on this point we refer the reader to [37]. We pass to the prime and dual problem (3.3) itself. For this we shall say that the family $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ is *eventually totally bounded in μ-mean* with respect to T , if the condition is satisfied (see [20] and [37]):

(3.7) For every $\varepsilon > 0$ there exists $\gamma_\varepsilon \in \Gamma(\Theta)$ such that:

$$\inf_{n \geq 1} \frac{1}{n} \int^* \sup_{\theta', \theta'' \in A} \Big| \sum_{j=0}^{n-1} T^j (f_{\theta'} - f_{\theta''}) \Big| \, d\mu < \varepsilon$$

for all $A \in \gamma_\varepsilon$.

Here and in the sequel $\Gamma(\Theta)$ denotes the family of all finite covers of Θ . We recall that a finite cover of Θ is any family $\gamma = \{ A_1, \ldots, A_n \}$ of non-empty subsets of Θ satisfying $\Theta = \cup_{j=1}^n A_j$ with $n \geq 1$. It is our next aim to show that under (3.1), condition (3.7) is equivalent to the uniform convergence in (3.3), with respect to any of the convergence notions stated above. It should be noted that our method below in essence relies upon the uniform ergodic lemma from the previous section. The first result may be stated as follows.

Theorem 3.1

Let $(X, \mathcal{B}, \mu, T)$ be a perfect ergodic dynamical system, and let $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ be a parameterized family of measurable maps from X into $\mathbf{R}$ satisfying (3.1). If $\mathcal{F}$ is eventually totally bounded in μ-mean with respect to T , then we have:

$$\sup_{\theta \in \Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \longrightarrow 0 \quad (a.s.)^* \ \& \ (L^1)^* \tag{3.8}$$

as $n \to \infty$ *, where* M *is the* μ*-mean function of* $\mathcal{F}$ *.*

Proof. Let $\varepsilon > 0$ be given and fixed, then by our assumption there exists $\gamma_\varepsilon \in \Gamma(\Theta)$ such that:

$$\inf_{n \geq 1} \int^* \sup_{\theta', \theta'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| \, d\mu < \varepsilon \tag{3.9}$$

for all $A \in \gamma_\varepsilon$. Since $M \in B(\Theta)$, then there is no restriction to assume that we also have:

$$\sup_{\theta', \theta'' \in A} |M(\theta') - M(\theta'')| < \varepsilon \tag{3.10}$$

for all $A \in \gamma_\varepsilon$. In addition choose a point $\theta_A \in A$ for every $A \in \gamma_\varepsilon$. Then from (3.10) we get:

$$\sup_{\theta \in \Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| = \max_{A \in \gamma_\varepsilon} \sup_{\theta \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \leq$$

$$\max_{A \in \gamma_\varepsilon} \Big(\sup_{\theta', \theta'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| + \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta_A}) - M(\theta_A) \Big| +$$

$$\sup_{\theta', \theta'' \in A} |M(\theta') - M(\theta'')| \Big) \leq \max_{A \in \gamma_\varepsilon} \sup_{\theta', \theta'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| +$$

$$\max_{A \in \gamma_\varepsilon} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta_A}) - M(\theta_A) \Big| + \varepsilon \ .$$

for all $n \geq 1$. Hence by (3.2) we easily obtain:

$$\limsup_{n \to \infty} \Big(\sup_{\theta \in \Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \Big)^* \leq$$

$$\max_{A \in \gamma_\varepsilon} \limsup_{n \to \infty} \Big(\sup_{\theta', \theta'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| \Big)^* + \varepsilon \quad \mu\text{-a.s.}$$

From this inequality and (3.9) we see that $(a.s.)^*$-convergence in (3.8) will be established as soon as we obtain the inequality:

$$\limsup_{n \to \infty} \Big(\sup_{\theta', \theta'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| \Big)^* \leq \inf_{n \geq 1} \int^* \sup_{\theta', \theta'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| \, d\mu \tag{3.11}$$

for all $A \subset \Theta$. We leave this inequality to be established with some facts of independent interest in the next proposition. We proceed with the $(L^1)^*$-convergence in (3.8). From (3.4) we get:

$$\Big(\sup_{\theta\in\Theta}\Big|\frac{1}{n}\sum_{j=0}^{n-1}T^j(f_\theta)-M(\theta)\Big|\Big)^* \leq \frac{1}{n}\sum_{j=0}^{n-1}\Big(\sup_{\theta\in\Theta}\big|T^j(f_\theta)\big|\Big)^* + \sup_{\theta\in\Theta}|M(\theta)| =$$
$$\frac{1}{n}\sum_{j=0}^{n-1}T^j\Big((\sup_{\theta\in\Theta}|f_\theta|)^*\Big) + \sup_{\theta\in\Theta}|M(\theta)|$$

for all $n \geq 1$. Hence we see that the sequence on the left-hand side is uniformly integrable. Thus $(L^1)^*$-convergence in (3.8) follows from $(a.s.)^*$-convergence, and the proof is complete. □

Proposition 3.2
Under the hypotheses of Theorem 3.1 we have:

(3.12) $$\limsup_{n\to\infty}\Big(\sup_{\theta',\theta''\in A}\Big|\frac{1}{n}\sum_{j=0}^{n-1}T^j(f_{\theta'}-f_{\theta''})\Big|\Big)^* = C \quad \mu\text{-a.s.}$$

(3.13) $$C \leq \inf_{n\geq 1}\int^* \sup_{\theta',\theta''\in A}\Big|\frac{1}{n}\sum_{j=0}^{n-1}T^j(f_{\theta'}-f_{\theta''})\Big|\, d\mu$$

(3.14) $$\inf_{n\geq 1}\int^* \sup_{\theta',\theta''\in A}\Big|\frac{1}{n}\sum_{j=0}^{n-1}T^j(f_{\theta'}-f_{\theta''})\Big|\, d\mu = \limsup_{n\to\infty}\int^* \sup_{\theta',\theta''\in A}\Big|\frac{1}{n}\sum_{j=0}^{n-1}T^j(f_{\theta'}-f_{\theta''})\Big|\, d\mu$$

where A is an arbitrary subset of Θ , and C is a real number depending on A .

First Proof: We could notice that if T is μ-perfect, then T^{j-1} is μ-perfect for all $j \geq 1$. This fact easily implies that the map $x \mapsto (T^0(x), T^1(x), \dots)$ is μ-perfect as a map from $(X, \mathcal{B})$ into $(X^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}})$. Hence we see that if T is assumed to be μ-perfect, then the hypotheses of Theorem 1 and Proposition 2 from Section 3 in [37] are satisfied with $\xi_j = T^{j-1}$ for $j \geq 1$. The proof then could be carried out in exactly the same way as the proof of Proposition 2 from Section 3 in [37]. This completes the first proof.

Second Proof: We could deduce the result of Proposition 3.2 by using the subadditive ergodic theorem of Kingman (see [12] p.292-299). Indeed, let us in the notation of Proposition 3.2 put:

$$g_n = \Big(\sup_{\theta',\theta''\in A}\Big|\sum_{j=0}^{n-1}T^j(f_{\theta'}-f_{\theta''})\Big|\Big)^*$$

for all $n \geq 1$, and let us assume that T is μ-perfect. Then it is quite easily verified that the sequence $\{ g_n \mid n \geq 1 \}$ is subadditive in $L^1(\mu)$. Therefore (3.12), (3.13) and (3.14) in Proposition 3.2 follow from Kingman's theorem in [12]. Moreover, even though it is irrelevant for our purposes, it should be noted that in this way we obtain immediately a more precise information. Namely, we have equality in (3.13) with the limit in (3.12) instead of the limit superior. (Since this follows from either the first proof or the third proof given below, together with the result of Theorem 3.1, we did not state the result of Proposition 3.2 in this stronger form. We have rather chosen the form which appears to be essential in the process of proving Theorem 3.1 by the method presented. For this one should recall (3.11) above.) Finally, it might be instructive in this context to recall the classical fact that a subadditive sequence of real numbers $\{ \gamma_n \mid n \geq 1 \}$ converges to $\inf_{n\geq 1}\gamma_n/n$ as $n \to \infty$. This completes the second proof.

Third Proof: We now give a proof of Proposition 3.2, based on our uniform ergodic lemma (see Theorem 2.1). Although this proof yields somewhat less (in comparison with the second proof), we include it for the sake of keeping the analogy with Birkhoff's theorem, with the hope that this method will find other applications.

Let $A \subset \Theta$ be given and fixed. Let us denote:

$$\Psi_n(x) \;=\; \sup_{\theta',\theta''\in A} \Big|\, \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''})(x) \,\Big|$$

for all $x \in X$ and all $n \geq 1$. Then the proof might be carried out as follows.

(3.12): Since T is ergodic, then it is enough to show that:

$$\limsup_{n\to\infty} \Psi_n^* \circ T \;=\; \limsup_{n\to\infty} \Psi_n^* \quad \mu\text{-a.s.} \tag{3.15}$$

For this it should be noted that we have:

$$\Psi_n \circ T \;=\; \sup_{\theta',\theta''\in A} \Big|\, \tfrac{n+1}{n} \cdot \tfrac{1}{n+1} \textstyle\sum_{j=0}^{n} T^j(f_{\theta'} - f_{\theta''}) - \tfrac{1}{n}\,(f_{\theta'} - f_{\theta''}) \,\Big|$$

for all $n \geq 1$. Hence we easily find:

$$\tfrac{n+1}{n}\, \Psi_{n+1} - \tfrac{2}{n} \sup_{\theta\in\Theta} |f_\theta| \leq \Psi_n \circ T \;\leq\; \tfrac{n+1}{n}\, \Psi_{n+1} + \tfrac{2}{n} \sup_{\theta\in\Theta} |f_\theta|$$

for all $n \geq 1$. Taking upper μ-envelopes on both sides we obtain:

$$\tfrac{n+1}{n}\, \Psi_{n+1}^* - \tfrac{2}{n} \big(\sup_{\theta\in\Theta} |f_\theta| \big)^* \leq \big(\Psi_n \circ T \big)^* \leq \tfrac{n+1}{n}\, \Psi_{n+1}^* + \tfrac{2}{n} \big(\sup_{\theta\in\Theta} |f_\theta| \big)^*$$

for all $n \geq 1$. Letting $n \to \infty$, and using (3.1)+(3.4) we obtain (3.15), and the proof of (3.12) is complete.

(3.13): Here we essentially use the uniform ergodic lemma from the last section. For this we denote $S_n(f) = \sum_{j=0}^{n-1} T^j f$ and $M_n(f) = \max_{1\leq j\leq n} S_j(f)$, whenever $f \in L^1(\mu)$ and $n \geq 1$. Let us for fixed $N, m \geq 1$ and $\varepsilon > 0$ consider a set as follows:

$$B_{N,m,\varepsilon} = \Big\{ \Big(\sup_{\theta',\theta''\in A} M_m\big(|S_N(f_{\theta'} - f_{\theta''})/N| - (C-\varepsilon) \big) \Big)^* > 0 \Big\} \;.$$

Then by Theorem 2.1 we may conclude:

$$\int_{B_{N,m,\varepsilon}}^* \sup_{\theta',\theta''\in A} \big(|S_N(f_{\theta'} - f_{\theta''})/N| - (C-\varepsilon) \big) \; d\mu \geq 0 \;.$$

Hence by subadditivity of upper μ-integral we obtain:

$$\int_{B_{N,m,\varepsilon}}^* \sup_{\theta',\theta''\in A} |S_N(f_{\theta'} - f_{\theta''})/N| \; d\mu \;\geq\; (C-\varepsilon)\, \mu(B_{N,m,\varepsilon}) \;.$$

Therefore by monotone convergence theorem for upper μ-integral (see [35]) in order to complete the proof of (3.13) it is enough to show that $\mu(B_{N,m,\varepsilon}) \uparrow 1$ as $m \to \infty$ with $N \geq 1$ and $\varepsilon > 0$ being fixed. We shall establish this fact by proving the following inequality:

$$(3.16)\qquad \Big(\sup_{\theta',\theta''\in A} |S_n(f_{\theta'} - f_{\theta''})/n| \Big)^* \leq \Big(\sup_{\theta',\theta''\in A} \sup_{m\geq 1} S_m\big(|S_N(f_{\theta'} - f_{\theta''})/N| \big) \Big)^* + \Big(\tfrac{1}{n} \cdot \sup_{\theta',\theta''\in A} \big| \textstyle\sum_{N[n/N]\leq j<n} T^j(f_{\theta'} - f_{\theta''}) \big| \Big)^*$$

being valid for all $n \geq N$. For this it should be noted that we have:

$$(3.17)\qquad |S_n(f_{\theta'} - f_{\theta''})/n| \leq |S_{N[n/N]}(f_{\theta'} - f_{\theta''})/n| + \tfrac{1}{n} \big| \textstyle\sum_{N[n/N]\leq j<n} T^j(f_{\theta'} - f_{\theta''}) \big|$$

being valid for all $n \geq N$ and all $\theta', \theta'' \in A$. Moreover, given $n \geq N$, taking $m \geq 1$ large enough we get:

$$S_m\big(|S_N(f_{\theta'} - f_{\theta''})|\big) = \big| \textstyle\sum_{j=0}^{N-1} T^j(f_{\theta'} - f_{\theta''})\big| + \big| \sum_{j=1}^{N} T^j(f_{\theta'} - f_{\theta''})\big| + \ldots + \big| \sum_{j=N}^{2N-1} T^j(f_{\theta'} - f_{\theta''})\big| + \ldots \geq \big| \sum_{j=0}^{N-1} T^j(f_{\theta'} - f_{\theta''})\big| + \big| \sum_{j=N}^{2N-1} T^j(f_{\theta'} - f_{\theta''})\big| + \ldots + \big| \sum_{j=N([n/N]-1)}^{N[n/N]-1} T^j(f_{\theta'} - f_{\theta''})\big| \geq |S_{N[n/N]}(f_{\theta'} - f_{\theta''})|$$

for all $\theta', \theta'' \in A$. Hence we obtain:

$$(3.18)\qquad |S_{N[n/N]}(f_{\theta'} - f_{\theta''})/n| \leq S_m\big(|S_N(f_{\theta'} - f_{\theta''})/n|\big) \leq S_m\big(|S_N(f_{\theta'} - f_{\theta''})/N|\big)$$

with $n, N, m \geq 1$ as above. Thus (3.16) follows straightforward by (3.17) and (3.18). In addition, for the last term in (3.16) we have:

$$\Big(\tfrac{1}{n} \cdot \sup_{\theta',\theta''\in A} \big| \textstyle\sum_{N[n/N]\leq j<n} T^j(f_{\theta'} - f_{\theta''}) \big| \Big)^* \leq \tfrac{2}{n} \textstyle\sum_{N[n/N]\leq j<n} T^j\big((\sup_{\theta\in A} |f_\theta|)^*\big) = 2\Big(\tfrac{1}{n} S_n\big((\sup_{\theta\in A} |f_\theta|)^*\big) - \tfrac{N[n/N]}{n} \cdot \tfrac{1}{N[n/N]} S_{N[n/N]}\big((\sup_{\theta\in A} |f_\theta|)^*\big)\Big)$$

for all $n \geq N$. Since $N[n/N]/n \to 1$ as $n \to \infty$, then by (3.1) and Birkhoff's theorem, the right-hand side tends to zero μ-a.s. as $n \to \infty$. Letting $n \to \infty$ in (3.16) therefore we obtain:

$$(3.19)\qquad \limsup_{n\to\infty} \Big(\sup_{\theta',\theta''\in A} |S_n(f_{\theta'} - f_{\theta''})/n| \Big)^* \leq \Big(\sup_{\theta',\theta''\in A} \sup_{m\geq 1} S_m\big(|S_N(f_{\theta'} - f_{\theta''})/N| \big) \Big)^*$$

for all $N \geq 1$. Finally, from (3.12) and (3.19) we get:

$$\lim_{m\to\infty} \mu(B_{N,m,\varepsilon}) = \lim_{m\to\infty} \mu^*\Big\{ \sup_{\theta',\theta''\in A} M_m\big(|S_N(f_{\theta'} - f_{\theta''})/N| \big) > (C-\varepsilon) \Big\} = \mu^*\Big\{ \sup_{\theta',\theta''\in A} \sup_{m\geq 1} S_m\big(|S_N(f_{\theta'} - f_{\theta''})/N| \big) > (C-\varepsilon) \Big\} = \mu\Big\{ \Big(\sup_{\theta',\theta''\in A} \sup_{m\geq 1} S_m\big(|S_N(f_{\theta'} - f_{\theta''})/N| \big) \Big)^* > (C-\varepsilon) \Big\} \geq$$

$$\mu\Big\{ \limsup_{n\to\infty} \Big(\sup_{\theta',\theta''\in A} |S_n(f_{\theta'} - f_{\theta''})/n| \Big)^* > (C-\varepsilon) \Big\} = 1$$

This fact completes the proof of (3.13).

(3.14): It should be noted that we have:

$$\Big(\sup_{\theta',\theta''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| \Big)^* \leq \frac{2}{n} \sum_{j=0}^{n-1} T^j \big((\sup_{\theta\in A} |f_\theta|)^* \big)$$

for all $n \geq 1$. Therefore by (3.1) it follows that the sequence on the left-hand side is uniformly integrable. Thus by Fatou's lemma we can conclude:

$$\limsup_{n\to\infty} \int^* \sup_{\theta',\theta''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big| \, d\mu \leq C \ .$$

Hence (3.14) follows straightforward by (3.13). This fact completes the proof. $\square$

Remark 3.3

Under the hypotheses of Theorem 3.1 and Proposition 3.2 it is easily verified that in the case where the map $x \mapsto \sup_{\theta',\theta''\in A} \big| \sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''})(x) \big|$ is μ-measurable as a map from X into $\mathbf{R}$ for all $n \geq 1$ and given $A \subset \Theta$, the assumption of μ-perfectness of T is not needed for their conclusions remain valid.

Remark 3.4

It should be noted that putting $n = 1$ in the definition of eventual total boundedness in μ-mean, we obtain an extension of the Blum-DeHardt sufficient condition, which is the best known for the uniform law of large numbers, at least in the independent case. There are examples showing that a parameterized family could be eventually totally bounded in μ-mean with respect to a measure preserving transformation, though it does not satisfy the Blum-DeHardt condition (see [11] p.41). It is not a surprise indeed, since we shall see in the next theorem that the eventual total boundedness in μ-mean property characterizes the uniform ergodic theorems, and thus the uniform laws of large numbers as well. In this way we may conclude that we have obtained a characterization of the uniform ergodic theorem which contains the best known sufficient condition as a particular case. Finally, let us mention a well-known and interesting fact that Blum-DeHardt's law of large numbers contains the classical Mourier's law of large numbers (see [11] p.43).

In order to state the next theorem we clarify that $B(\Theta)$ denotes the Banach space of all bounded real valued functions on Θ with respect to the sup-norm. We denote by $C(B(\Theta))$ the set of all bounded continuous functions from $B(\Theta)$ into $\mathbf{R}$, while by $\mathcal{K}(B(\Theta))$ we denote the family of all compact subsets of $B(\Theta)$. A pseudo-metric d on a set Θ is said to be *totally bounded*, if Θ can be covered by finitely many d-balls of any given radius $\varepsilon > 0$. A pseudo-metric d on a set Θ is said to be a *ultra pseudo-metric*, if it satisfies $d(\theta_1,\theta_2) \leq d(\theta_1,\theta_3) \vee d(\theta_3,\theta_2)$ for all $\theta_1,\theta_2,\theta_3 \in \Theta$.

Theorem 3.5

Let $(X,\mathcal{B},\mu,T)$ *be a perfect ergodic dynamical system, and let* $\mathcal{F} = \{\, f_\theta \mid \theta \in \Theta \,\}$ *be a parameterized family of measurable maps from* X *into* $\mathbf{R}$ *satisfying (3.1). Let* M *be the* μ*-mean* *function of* $\mathcal{F}$ *, then the following eight statements are equivalent:*

(3.20) *The family* $\mathcal{F}$ *is eventually totally bounded in* μ*-mean with respect to* T

(3.21) $$\sup_{\theta\in\Theta} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \longrightarrow 0 \;\; (a.s.)^*$$

(3.22) $$\sup_{\theta\in\Theta} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \longrightarrow 0 \;\; (L^1)^*$$

(3.23) $$\sup_{\theta\in\Theta} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \Big| \longrightarrow 0 \;\; (\mu^*)$$

(3.24) $\frac{1}{n}\sum_{j=0}^{n-1} T^j(f) \rightarrow M$ *weakly in* $B(\Theta)$ *:*

$$\lim_{n\to\infty} \int^* F\Big(\frac{1}{n}\sum_{j=0}^{n-1} T^j(f)\Big)\, d\mu = F(M)$$

for all $F \in C(B(\Theta))$

(3.25) *The sequence* $\Big\{ \frac{1}{n}\sum_{j=0}^{n-1} T^j(f) \mid n \geq 1 \Big\}$ *is eventually tight in* $B(\Theta)$ *:*

$$\limsup_{n\to\infty} \int^* F\Big(\frac{1}{n}\sum_{j=0}^{n-1} T^j(f)\Big)\, d\mu \leq \varepsilon$$

for some $K_\varepsilon \in \mathcal{K}(B(\Theta))$ *, and for all* $F \in C(B(\Theta))$ *satisfying* $0 \leq F \leq 1_{B(\Theta)\setminus K_\varepsilon}$ *, whenever* $\varepsilon > 0$

(3.26) *There exists a totally bounded ultra pseudo-metric* d *on* Θ *such that the condition is satisfied:*

$$\lim_{r\downarrow 0}\, \inf_{n\geq 1} \int^* \sup_{d(\theta',\theta'')<r} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_{\theta'} - f_{\theta''}) \Big|\, d\mu = 0$$

(3.27) *For every* $\varepsilon > 0$ *there exist a totally bounded pseudo-metric* d_ε *on* Θ *and* $r_\varepsilon > 0$ *such that:*

$$\inf_{n\geq 1} \mu^*\Big\{ \sup_{d_\varepsilon(\theta,\theta')<\varepsilon} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\theta - f_{\theta'}) \Big| > \varepsilon \Big\} < \varepsilon$$

for all $\theta \in \Theta$.

Proof. Since T is μ-perfect, then T^{j-1} is μ-perfect for all $j \geq 1$. This fact as before easily implies that the map $x \mapsto (T^0(x), T^1(x), \dots)$ is μ-perfect as a map from $(X, \mathcal{B})$ into $(X^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}})$. Putting $\xi_j = T^{j-1}$ for $j \geq 1$ we see that the hypotheses of Theorem 4, Corollary 5, Theorem 7, Corollary 8 and Theorem 10 from Section 3 in [37] are satisfied. The result therefore follows from Corollary 5, Theorem 7, Corollary 8 and Theorem 10 from Section 3 in [37]. It should be noted that a characterization of compact sets in the Banach space $B(\Theta)$ is used (see

[13] p.260) for sufficiency of (3.25). It is also not difficult to verify (by using uniform integrability) that (3.23) implies (3.20). These facts complete the proof. $\square$

4. Examples and complements

We continue as in the preceding section by considering a given dynamical system $(X, \mathcal{B}, \mu, T)$ and a parameterized family $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ of measurable maps from X into $\mathbf{R}$. First we discuss a necessary condition for the prime and dual problem as follows. Suppose that (3.3) is satisfied with respect to $(a.s.)^*$-convergence. We could then question does this fact imply validity of condition (3.1). The answer is negative in general, and a more detailed discussion on this point with references can be found in [37]. Here we show that a weaker condition still remains true. The result may be stated as follows.

Proposition 4.1

Let $(X, \mathcal{B}, \mu, T)$ be a perfect dynamical system, and let $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ be a parameterized family of measurable maps from X into $\mathbf{R}$. Suppose that:

$$\limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) \Big| \Big)^* < \infty \quad \mu\text{-}a.s. \tag{4.1}$$

Then we have:

$$\Big(\sup_{\theta\in\Theta} |f_\theta| \Big)^* < \infty \quad \mu\text{-}a.s. \tag{4.2}$$

Proof. We show that under the primary hypotheses stated above we have:

$$\mu\Big\{ \limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) \Big| \Big)^* < \infty \Big\} \leq \mu\Big\{ \big(\sup_{\theta\in\Theta} |f_\theta| \big)^* < \infty \Big\} . \tag{4.3}$$

For this it should be noted that we have:

$$T^n(f_\theta) = (n+1) \cdot \tfrac{1}{n+1} \textstyle\sum_{j=0}^{n} T^j(f_\theta) - n \cdot \tfrac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta)$$

for all $n \geq 1$ and all $\theta \in \Theta$. Hence we get:

$$\big(\sup_{\theta\in\Theta} |T^n(f_\theta)| \big)^* \leq (n+1) \cdot \Big(\sup_{\theta\in\Theta} \big| \tfrac{1}{n+1} \textstyle\sum_{j=0}^{n} T^j(f_\theta) \big| \Big)^* +$$

$$n \cdot \Big(\sup_{\theta\in\Theta} \big| \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(f_\theta) \big| \Big)^*$$

for all $n \geq 1$. From this and the μ-perfectness of T we may easily conclude:

$$\mu\Big\{ \limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) \Big| \Big)^* < \infty \Big\} =$$

$$\mu\Big(\bigcup_{M=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{ \Big(\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) \Big| \Big)^* \leq M \Big\} \Big) \leq$$

$$\mu\Big(\bigcup_{M=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{ \Big(\sup_{\theta\in\Theta} \mid T^n(f_\theta) \mid \Big)^* \leq \ (2n+1)\, M \Big\} \Big) \leq$$

$$\mu\Big(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{ \Big(\sup_{\theta\in\Theta} \mid T^n(f_\theta) \mid \Big)^* < \infty \Big\} \Big) \leq$$

$$\liminf_{n\to\infty} \ \mu\Big\{ \Big(\sup_{\theta\in\Theta} \mid T^n(f_\theta) \mid \Big)^* < \infty \Big\} = \mu\Big\{ \ \big(\sup_{\theta\in\Theta} |f_\theta| \big)^* < \infty \Big\} \ .$$

Actually, in the last step we use the fact that T^n is μ-perfect for every $n \geq 1$. Thus (4.3) is satisfied, and (4.2) follows straightforward by (4.1). These facts complete the proof. $\square$

Corollary 4.2

Under the hypotheses of Proposition 4.1 suppose moreover that $\mathcal{F} \subset L^1(\mu)$ *and that (3.3) is satisfied with respect to either* $(a.s.)^*$*-convergence,* $(L^1)^*$*-convergence, or* (μ^*)*-convergence, where* M *is the* μ*-mean function of* $\mathcal{F}$ *satisfying* $\sup_{\theta\in\Theta}|M(\theta)| < \infty$ *. Then (4.2) in Proposition 4.1 is satisfied.*

Proof. It follows easily from Proposition 4.1 by using the inequality:

$$\limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \mid \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(f_\theta) \mid \Big)^* \leq$$

$$\limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \mid \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(f_\theta) - M(\theta) \mid \Big)^* + \ \sup_{\theta\in\Theta}|M(\theta)| \ .$$

This completes the proof. $\square$

Next we investigate a uniform approximation in the L^1-sense by means of a dense family from $L^1(\mu)$ which satisfies the uniform ergodic theorem in a trivial way. This approach is in the spirit of the classical Hilbert space method for the mean ergodic theorem of von Neumann, and therefore from the ergodic theory point of view it could be seen as the natural one. We consider as before a given ergodic dynamical system $(X, \mathcal{B}, \mu, T)$ and a parameterized family $\mathcal{F} = \{\, f_\theta \mid \theta \in \Theta \,\}$ of measurable maps from X into $\mathbf{R}$. We recall that $L^2(\mu)$ is a Hilbert space which is dense in $L^1(\mu)$, and for which by the general Hilbert space theory we have (see [25] p.4):

$$L^2(\mu) \ = \ G \oplus cl(H)$$

where $G = \{\, g \in L^2(\mu) \mid Tg = g \,\}$ and $H = \{\, h - Th \mid h \in L^\infty(\mu) \,\}$. It is instructive to observe in this context that G consists of constants, since T is assumed to be ergodic. In addition, suppose that maps $g_\theta = h_\theta - Th_\theta \in H$ for $\theta \in \Theta$ are given, such that $\big(\sup_{\theta\in\Theta} |h_\theta| \big)^* \in L^1(\mu)$. Then by Birkhoff's pointwise ergodic theorem we have:

$$\text{(4.4)} \qquad \Big(\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(g_\theta) \Big| \Big)^* \ = \ \Big(\sup_{\theta\in\Theta} \Big| \frac{1}{n} \big(h_\theta - T^n(h_\theta) \big) \Big| \Big)^*$$

$$\leq \ \frac{1}{n} \Big(\sup_{\theta\in\Theta} |h_\theta| \Big)^* + \ \frac{1}{n} \ T^n \Big(\big(\sup_{\theta\in\Theta} |h_\theta| \,\big)^* \Big) \ \to 0 \quad \mu\text{-a.s.}$$

as $n \to \infty$. In this way we obtain a natural family of maps satisfying the uniform ergodic theorem in a trivial way. Moreover, let us put $\Phi^\infty(H)$ to denote the class of all families $Z = \{ h_\theta - Th_\theta \mid \theta \in \Theta \} \subset H$ with $\big(\sup_{\theta\in\Theta} |h_\theta|\,\big)^* \in L^1(\mu)$. Then every g_θ from $Z \in \Phi^\infty(H)$ for $\theta \in \Theta$ satisfies (4.4), and the class $\Phi^\infty(H)$ could be enlarged to satisfy the uniform ergodic theorem as follows.

Theorem 4.3

Let $(X, \mathcal{B}, \mu, T)$ be a perfect ergodic dynamical system, and let $\mathcal{F} = \{ f_\theta \mid \theta \in \Theta \}$ be a parameterized family of measurable maps from X into $\mathbf{R}$ satisfying $\int f_\theta \, d\mu = 0$ for all $\theta \in \Theta$. Suppose that the following condition is satisfied:

(4.5) *For every $\varepsilon > 0$, there exists $Z_\varepsilon = \{ g_{\theta,\varepsilon} \mid \theta \in \Theta \} \in \Phi^\infty(H)$ satisfying:*

$$\int^* \sup_{\theta\in\Theta} |f_\theta - g_{\theta,\varepsilon}| \, d\mu < \varepsilon \, .$$

Then we have:

(4.6)
$$\sup_{\theta\in\Theta} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\theta) \Big| \longrightarrow 0 \quad (a.s.)^* \ \& \ (L^1)^*$$

as $n \to \infty$.

Proof. Let $\varepsilon > 0$ be given, then there exists $Z_\varepsilon = \{ g_{\theta,\varepsilon} \mid \theta \in \Theta \} \in \Phi^\infty(H)$ satisfying the inequality in (4.5). Hence by Birkhoff's theorem and (4.4) we get:

$$\limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \big| \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(f_\theta) \big| \Big)^* \leq$$
$$\limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \big| \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(f_\theta - g_{\theta,\varepsilon}) \big| \Big)^* +$$
$$\limsup_{n\to\infty} \Big(\sup_{\theta\in\Theta} \big| \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(g_{\theta,\varepsilon}) \big| \Big)^* \leq$$
$$\lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j\big((\sup_{\theta\in\Theta} |f_\theta - g_{\theta,\varepsilon}|)^*\big) = \int^* \sup_{\theta\in\Theta} |f_\theta - g_{\theta,\varepsilon}| \, d\mu < \varepsilon \, .$$

This fact establishes $(a.s.)^*$-convergence in (4.6). Since $\big(\sup_{\theta\in\Theta} |g_{\theta,1}|\,\big)^* \in L^1(\mu)$, then it is easily seen from (4.5) (with $\varepsilon = 1$) that condition (3.1) is fulfilled. Since we have:

$$\Big(\sup_{\theta\in\Theta} \big| \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j(f_\theta) \big| \Big)^* \leq \tfrac{1}{n} \textstyle\sum_{j=0}^{n-1} T^j\big((\sup_{\theta\in\Theta} |f_\theta|)^*\big)$$

for all $n \geq 1$, then by (3.1) we see that $(L^1)^*$-convergence in (4.6) follows from (4.5) and $(a.s.)^*$-convergence in (4.6) by uniform integrability. This fact completes the proof. □

We pass to a question of independent interest. It concerns a uniform convergence of *moving averages*, and its characterization obtained by means of *Banach limits*. We recall that a linear functional L on l_∞ is called a *Banach limit*, if the following three conditions are satisfied:

(4.7) $L(x) \geq 0$ whenever $x \geq 0$, meaning that all coordinates of x are non-negative

(4.8) $\qquad L(x_1, x_2, x_3, \dots) = L(x_2, x_3, \dots)$ for all $(x_1, x_2, x_3, \dots) \in l_\infty$

(4.9) $\qquad L(\mathbf{1}) = 1$ where $\mathbf{1} = (1, 1, 1, \dots)$.

It is well-known that Banach limits exist. Moreover if L is a Banach limit, then we have:

(4.10) $\qquad \| L \| = 1$

(4.11) $\qquad \liminf_{n\to\infty} x_n \le L(x) \le \limsup_{n\to\infty} x_n$.

In addition we have:

(4.12) $\qquad \sup \{ L(x) \mid L \text{ is a Banach limit} \} = \lim_{n\to\infty} \sup_{i\ge1} \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i}$

for all $x = (x_1, x_2, \dots) \in l_\infty$. Finally, if $x = (x_1, x_2, \dots) \in l_\infty$ and $m \in \mathbf{R}$, then $L(x) = m$ for every Banach limit L , if and only if we have:

(4.13) $$\lim_{n\to\infty} \sup_{i\ge1} \left| \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i} - m \right| = 0 .$$

For proof of all these facts we shall refer the reader to [25] (p.135-136). However, the last one seems to be a weak characterization of (4.13), since there could be too many Banach limits for verification. A more convenient characterization of (4.13) may be stated as follows.

Proposition 4.4

Let $x = (x_1, x_2, \dots)$ be an arbitrary element of l_∞ , and let m be a real number. Then (4.13) is valid, if and only if the following two conditions are satisfied:

(4.14) $\qquad \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} x_j = m$

(4.15) $\qquad$ *The set* $\left\{ \left(n^{-1} \sum_{j=0}^{n-1} x_{j+i} \right)_{n\ge1} \mid i \ge 1 \right\}$ *is totally bounded in* l_∞ .

Proof. It follows easily by definition that (4.13) implies (4.14) and (4.15). Conversely, it should be noted that (4.14) is equivalent to:

(4.16) $$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i} = m$$

being valid for all $i \ge 1$. In addition, let $\varepsilon > 0$ be given and fixed. Then by (4.15) there exist $i_1, \dots, i_p \ge 1$ such that:

$$\min_{1\le k\le p} \sup_{n\ge1} \left| \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i} - \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i_k} \right| < \varepsilon$$

for all $i \ge 1$. Moreover, by (4.16) we can choose $n \ge N_\varepsilon$ large enough to have:

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i_k} - m \right| < \varepsilon$$

for all $k = 1, \dots, p$. Hence by the triangle inequality we easily obtain:

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} x_{j+i} - m \right| < 2\varepsilon$$

for all $n \geq N_\varepsilon$ and all $i \geq 1$. This fact completes the proof. $\square$

Remark 4.5

One could think that condition (4.15) in Proposition 4.4 may be replaced by the following condition:

(4.15') The set $\{ (x_{n-1+i})_{n\geq 1} \mid i \geq 1 \}$ is totally bounded in l_∞ .

However, this is not true. Indeed, even though (4.15') implies (4.15), and therefore (4.14)+(4.15') implies (4.13), we could have (4.13) without (4.15') being valid. For example, take x to be $(1, -1, 0, 1, -1, 0, 0, 1, -1, 0, 0, 0, 1, -1, \ldots)$. Then (4.13) holds with $m = 0$, but we have $\| (x_{n-1+i_1})_{n\geq 1} - (x_{n-1+i_2})_{n\geq 1} \|_\infty \geq 1$ whenever $i_1 \neq i_2$.

In the next example we show that replacing the unit ball of a Banach space in the prime problem from Section 1 by a smaller family of vectors allows the uniform ergodic theorem (1.1) to become weaker and more easily established (but still of a considerable interest). Our emphasis is rather on the interpretation of the result than on the generality in which it is stated.

Example 4.6 (**A uniform ergodic theorem over the orbit**)

Consider the d-dimensional torus $X = [0,1[^d$ equipped with the d-dimensional Lebesgue measure $\mu = \lambda_d$ for some $d \geq 1$. Then X is a compact group with respect to the coordinatewise addition mod 1. Take a point $\alpha = (\alpha_1, \ldots, \alpha_d)$ in X such that $\alpha_1, \ldots, \alpha_d$ and 1 are integrally independent (if $\sum_{i=1}^d k_i \, \alpha_i$ is an integer for some integers $k_1, \ldots, k_d$, then $k_1 = \ldots = k_d = 0$). Under this condition the translation $T(x) = x + \alpha$ is an ergodic measure-preserving transformation of X , and the orbit $\mathcal{O}(x) = \{ T^j(x) \mid j \geq 0 \}$ is dense in X for every $x \in X$ (see [25] p.12). Let $f : X \to \mathbf{R}$ be a continuous function, and let $\mathcal{O}(f) = \{ T^j(f) \mid j \geq 0 \}$ be the orbit of f under T . Then the prime problem from Section 1 may be restated to a problem of the uniform convergence as follows:

$$\sup_{g \in \mathcal{O}(f)} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(g) - \int f \, d\mu \Big| \longrightarrow 0 \tag{4.17}$$

as $n \to \infty$. It is immediate that (4.17) is equivalent to the following statement:

$$\sup_{i \geq 0} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^{i+j}(f) - \int f \, d\mu \Big| \longrightarrow 0 \tag{4.18}$$

as $n \to \infty$. Finally, we obtain by the continuity of f and denseness of $\mathcal{O}(x)$ for every $x \in X$, that (4.18) is equivalent to the following statement:

$$\sup_{x \in X} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f)(x) - \int f \, d\mu \Big| \longrightarrow 0 \tag{4.19}$$

as $n \to \infty$. However, this is precisely the statement of *Weyl's theorem* on uniform distribution mod 1 (see [25] p.13). This establishes the uniform ergodic theorem (4.17).

In fact, the equivalence of (4.17)-(4.19) extends to any Markov operator T in $C(X)$ with

X being a compact Hausdorff space. For this, it should be noted that the pointwise convergence in (4.18) with $i = 0$ means weak convergence in $C(X)$, and implies the strong convergence (see [25] p.72) which is precisely (4.19). For more information in this direction see [25] (p.12) and [40] (p.137).

So far we have considered a Blum-DeHardt approach towards solution for the prime and dual problem by offering Lipschitz type conditions. The next lines concern the Vapnik-Chervonenkis random entropy approach. We begin in this direction by displaying a cornerstone for this type of results in the independent case. For this consider a sequence of independent and identically distributed random variables $\{\xi_j \mid j \geq 1\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$. Let $f : S \times \Theta \to \mathbf{R}$ be a given map, such that $M(\theta) = \int f(\xi_1, \theta)\, dP$ is well-defined. Let $\{\varepsilon_j \mid j \geq 1\}$ be a Bernoulli sequence independent of $\{\xi_j \mid j \geq 1\}$. Then passing over measurability problems it is well-known that we have:

$$E \sup_{\theta \in \Theta} \Big| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j, \theta) - M(\theta) \Big| \leq 2 \cdot E \sup_{\theta \in \Theta} \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, \theta) \Big| \tag{4.20}$$

for all $n \geq 1$. It turns out that this symmetrization inequality plays a vitally important role in the Vapnik-Chervonenkis random entropy approach. Moreover, it might be easily verified that having this inequality in the case where the sequence $\{\xi_j \mid j \geq 1\}$ is only assumed to be stationary and ergodic, the basic random entropy result of Vapnik and Chervonenkis could be established in exactly the same way as in the independent case (see for instance the proof in [53]). However, our next example shows that this is not possible in general.

Example 4.7

We show that inequality (4.20) may fail if the sequence $\{\xi_j \mid j \geq 1\}$ is only assumed to be stationary and ergodic. For this we shall consider a simple case where Θ consists of a single point, and where f equals to the identity map on the real line. The sequence $\{\xi_j \mid j \geq 1\}$ itself is for a moment only assumed to be stationary and Gaussian. Thus our question reduces to verify the following inequality:

$$E \Big| \frac{1}{n} \sum_{j=1}^{n} \xi_j \Big| \leq C \cdot E \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot \xi_j \Big| \tag{4.21}$$

for all $n \geq 1$ with some constant $C > 0$, where $\{\varepsilon_j \mid j \geq 1\}$ is a Bernoulli sequence independent of $\{\xi_j \mid j \geq 1\}$. Let us first consider the right-hand side of this inequality. For this denote by $\|\cdot\|_{\Psi_2}$ the Orlicz norm induced by the function $\Psi_2(x) = \exp(x^2) - 1$ for $x \in \mathbf{R}$. Then it is easily verified that we have $\|X\|_1 \leq 6/5\, \|X\|_{\Psi_2}$ whenever X is a random variable. Hence by Kahane-Khintchine's inequality for $\|\cdot\|_{\Psi_2}$ and Jensen's inequality we get (see [38]):

$$\begin{aligned} E \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot \xi_j \Big| &= E_\xi\, E_\varepsilon \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot \xi_j \Big| \leq \frac{6}{5} \cdot \frac{1}{n} \cdot E_\xi \Big\| \sum_{j=1}^{n} \varepsilon_j \cdot \xi_j \Big\|_{\psi_2, \varepsilon} \leq \\ &\frac{6}{5} \cdot \sqrt{\frac{8}{3}} \cdot \frac{1}{n} \cdot E_\xi \Big(\sum_{j=1}^{n} |\xi_j|^2 \Big)^{1/2} \leq 2 \cdot \frac{1}{n} \sqrt{n\, E|\xi_1|^2} = \frac{2}{\sqrt{n}} \cdot \sqrt{E|\xi_1|^2} \end{aligned} \tag{4.22}$$

for all $n \geq 1$. On the other hand since $n^{-1} \sum_{j=1}^{n} \xi_j$ is Gaussian, then we have:

$$\Big(E \, \Big| \, \frac{1}{n} \sum_{j=1}^{n} \xi_j \, \Big|^2 \, \Big)^{1/2} \leq \; D \cdot E \, \Big| \, \frac{1}{n} \sum_{j=1}^{n} \xi_j \, \Big| \tag{4.23}$$

for some constant $D > 0$ and all $n \geq 1$. Inserting (4.22) and (4.23) into (4.21) we obtain:

$$E \, \Big| \sum_{j=1}^{n} \xi_j \Big|^2 \leq G \cdot n \cdot E|\xi_1|^2 \tag{4.24}$$

for all $n \geq 1$ with $G = \sqrt{2CD}$. Thus it is enough to show that (4.24) may fail in general. Since $\{ \xi_j \mid j \geq 1 \}$ is stationary, then we have $E(\xi_i \xi_j) = R(i-j)$ for all $i, j \geq 1$. Moreover, it is easily verified that the left-hand side in (4.24) may be written as follows:

$$E \, \Big| \sum_{j=1}^{n} \xi_j \Big|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} E(\xi_i \xi_j) \; = \; \sum_{i=1}^{n} \sum_{j=1}^{n} R(i-j) \; = \; n \, R(0) + \sum_{k=1}^{n-1} 2 \, (n-k) \, R(k) \tag{4.25}$$

for all $n \geq 1$. Let us in addition consider a particular case by putting $R(k) = 1/(k+1)$ for all $k \geq 0$. Then R is a decreasing convex function, and therefore by Polya's theorem it could be the covariance function of a centered stationary Gaussian sequence $\{ \xi_j \mid j \geq 1 \}$. Moreover, by Maruyama's theorem this sequence is strongly mixing, and thus ergodic as well. Finally, from (4.25) we may easily obtain:

$$\begin{aligned} & E \, \Big| \sum_{j=1}^{n} \xi_j \Big|^2 = \; n + 2 \sum_{k=1}^{n-1} \frac{n-k}{k+1} \; = \; 2 \, (n+1) \sum_{k=1}^{n-1} \frac{1}{k+1} - n + 2 \geq \\ & 2 \, (n+1) \int_1^n \frac{1}{x+1} \; dx - n + 2 \; = \\ & 2 \, (n+1) \log{(n+1)} - (1 + \log 4) \, (n+1) + 3 \geq \; n \log n \end{aligned}$$

for all $n \geq 1$. This inequality contradicts (4.24), and therefore (4.21) is false in this case as well.

Remark 4.8

It could be instructive to observe that in order that inequality (4.24) from Example 4.7 holds we should have $\| \sum_{j=1}^{n} \xi_j \|_2 = O(\sqrt{n})$ as $n \rightarrow \infty$. This fact is known to be valid if the sequence $\{ \xi_j \mid j \geq 1 \}$ is mixing enough (see for instance [8]). It could indicate that the symmetrization method is intimately related with a mixing property. Moreover, even though Example 4.7 shows that the symmetrization inequality (4.20) does not extend to the stationary ergodic case, it does not contradict the most useful consequence of (4.20) in this context. Namely, for Vapnik-Chervonenkis type results in the notation of (4.20), it seems quite enough to have "an asymptotic symmetrization inequality" as follows:

$$E \sup_{\theta \in \Theta} \Big| \, \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, \theta) \, \Big| \rightarrow 0 \;\; \Rightarrow \;\; E \sup_{\theta \in \Theta} \Big| \, \frac{1}{n} \sum_{j=1}^{n} f(\xi_j, \theta) - M(\theta) \, \Big| \rightarrow 0 \tag{4.26}$$

as $n \rightarrow \infty$. Our Example 4.7 shows that these convergencies could have different speeds. These observations may leave some open space for further developments (see [39]).

Acknowledgment. The authors thank J. Hoffmann-Jørgensen and S. E. Graversen for useful discussions.

REFERENCES

[1] ATALLA, R. E. (1984). On uniform ergodic theorems for Markov operators on $C(X)$. *Rocky Mt. J. Math.* 14 (451-456).

[2] ATALLA, R. E. (1985). Correction to "On uniform ergodic theorems for Markov operators on $C(X)$ " . *Rocky Mt. J. Math.* 15 (763).

[3] BLUM, J. R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* 26 (527-529).

[4] BRUNEL, A. *and* REVUZ, D. (1974). Quelques applications probabilistes de la quasi-compacité. *Ann. Inst. H. Poincaré Probab. Statist.* 10 (301-337).

[5] BURKE, G. (1965). A uniform ergodic theorem. *Ann. Math. Statist.* 36 (1853-1858).

[6] CANTELLI, F. P. (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari* 4 (421-424).

[7] DEHARDT, J. (1971). Generalizations of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* 42 (2050-2055).

[8] DEHLING, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 63 (393-432).

[9] DOEBLIN, W. (1937/38). Sur les propriétés asymptotiques de mouvements régis par certains types de chaines simples. *Bull. Soc. Math. Roumaine* 39-1 (1937) (57-115) and 39-2 (1938) (3-61).

[10] DOOB, J. L. (1938). Stochastic processes with an integral valued parameter. *Trans. Amer. Math. Soc.* 44 (87-150).

[11] DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Eté de Probabilitiés de Saint-Flour, XII-1982. Lecture Notes in Math.* 1097, Springer-Verlag Berlin Heidelberg (1-142).

[12] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth, Inc., Belmont, California 94002.

[13] DUNFORD, N. *and* SCHWARTZ, J. T. (1958). *Linear Operators, Part I: General Theory*. Interscience Publ. Inc., New York.

[14] EBERLEIN, W. F. (1949). Abstract ergodic theorems and weak almost periodic functions. *Trans. Amer. Math. Soc.* 67 (217-240).

[15] GAENSSLER, P. (1983). Empirical Processes. *IMS Lecture Notes-Monograph Series* 3.

[16] GARSIA, A. M. (1970). Topics in Almost Everywhere Convergence. *Lectures in Advanced Mathematics* 4, Markham Publishing Company.

[17] GINÉ, J. *and* ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* 12 (929-989).

[18] GLIVENKO, V. I. (1933). Sulla determizaione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari* 4 (92-99).

[19] GROCH, U. (1984). Uniform ergodic theorems for identity preserving Schwarz maps on W^*-algebras. *J. Operator Theory* 11 (395-404).

[20] HOFFMANN-JØRGENSEN, J. (1984). Necessary and sufficient conditions for the uniform law of large numbers. *Proc. Probab. Banach Spaces* V (Medford 1984), *Lecture Notes in Math.* 1153, Springer-Verlag Berlin Heidelberg (258-272).

[21] HOFFMANN-JØRGENSEN, J. (1990). Uniform convergence of martingales. *Proc. Probab. Banach Spaces* VII (Oberwolfach 1988), *Progr. Probab.* Vol. 21, Birkhäuser Boston, (127-137).

[22] HOFFMANN-JØRGENSEN, J. (1991). The general monotone convergence theorem and measurable selections. *Math. Inst. Aarhus, Preprint Ser.* No. 13, (37 pp).

[23] HOROWITZ, S. (1972). Transition probabilities and contractions of L_∞ . *Z. Wahrscheinlichkeitstheorie verw. Geb.* 24 (263-274).

[24] JAJTE, R. (1984). A few remarks on the almost uniform ergodic theorems in von Neumann algebras. *Probability theory and vector spaces III, Proc. Conf. Lublin, Poland 1983, Lecture Notes in Math.* 1080, Springer-Verlag Berlin Heidelberg (120-143).

[25] KRENGEL, U. (1985). *Ergodic Theorems*. Walter de Gruyter & Co., Berlin.

[26] KRYLOFF, N. *and* BOGOLIOÙBOFF, N. (1937). Sur les propriétés en chaine. *C. R. Paris* 204 (1386-1388).

[27] KRYLOFF, N. *and* BOGOLIOÙBOFF, N. (1937). Les propriétés ergodiques des suites des probabilitiés en chaine. *C. R. Paris* 204 (1454-1456).

[28] LADOUCEUR, S. *and* WEBER, M. (1992). Speed of convergence of the mean average operator for quasi-compact operators. *Preprint*.

[29] LEDOUX, M. *and* TALAGRAND, M. (1991). *Probability in Banach Spaces* (*Isoperimetry and Processes*). Springer-Verlag Berlin Heidelberg.

[30] LIN, M. (1974). On the uniform ergodic theorem. *Proc. Amer. Math. Soc.* 43 (337-340).

[31] LIN, M. (1974). On the uniform ergodic theorem II. *Proc. Amer. Math. Soc.* 46 (217- 225).

[32] LIN, M. (1978). Quasi-compactness and uniform ergodicity of positive operators. *Israel J. Math.* 29 (309-311).

[33] LLOYD, S. P. (1981). On the uniform ergodic theorem of Lin. *Proc. Amer. Math. Soc.* 83 (710-714).

[34] LOTZ, H. P. (1981). Uniform ergodic theorems for Markov operators on $C(X)$. *Math. Z.* 178 (145-156).

[35] PESKIR, G. (1991). Perfect measures and maps. *Math. Inst. Aarhus, Preprint Ser.* No. 26, (34 pp).

[36] PESKIR, G. (1992). Uniform convergence of reversed martingales. *Math. Inst. Aarhus, Preprint Ser.* No. 21, (27 pp). *J. Theoret. Probab.* 8, 1995 (387-415).

[37] PESKIR, G. *and* WEBER, M. (1992). Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Math. Inst. Aarhus, Preprint Ser.* No. 27, (26 pp). *Proc. Funct. Anal.* IV (Dubrovnik 1993), *Various Publ. Ser.* Vol. 43, 1994 (165-190).

[38] PESKIR, G. (1992). Best constants in Kahane-Khintchine inequalities in Orlicz spaces. *Math. Inst. Aarhus, Preprint Ser.* No. 10, (42 pp). *J. Multivariate Anal.* 45, 1993 (183-216).

[39] PESKIR, G. *and* YUKICH, J. E. (1993). Uniform ergodic theorems for measurable dynamical systems under VC entropy conditions. *Math. Inst. Aarhus, Preprint Ser.* No. 15, (25 pp). *Proc. Probab. Banach Spaces* IX (Sandbjerg 1993), *Progr. Probab.* Vol. 35, Birhauser, Boston, 1994 (104-127).

[40] PETERSEN, K. (1983). *Ergodic Theory*. Cambridge University Press.

[41] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag New York Inc.

[42] POLLARD, D. (1990). Empirical Processes: Theory and Applications. *NSF-CBMS Regional Conference Series in Probability in Statistics*, Vol. 2.

[43] SHAW, S. Y. (1986). Uniform ergodic theorems for locally integrable semigroups and pseudo-resolvents. *Proc. Amer. Math. Soc.* 98 (61-67).

[44] TALAGRAND, M. (1987). The Glivenko-Cantelli problem. *Ann. Probab.* 15 (837-870).

[45] VAPNIK, V. N. *and* CHERVONENKIS, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* 16 (264-280).

[46] VAPNIK, V. N. *and* CHERVONENKIS, A. Ya. (1981). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.* 26 (532-553).

[47] WEBER, M. (1990). Une version fonctionnelle du théorème ergodique poncuel. *C. R. Acad. Sci. Paris, Sér I Math.* 311 (131-133).

[48] WEBER, M. (1993). GC sets, Stein's elements and matrix summation methods. *IRMA, Strasbourg, Prépubl. No. 27.*

[49] WEBER, M. (1996). Coupling of the GB set property and uniformity for ergodic averages. *J. Theoret. Probab.* 9 (105-112).

[50] YOSIDA, K. *and* KAKUTANI, S. (1938). Application of mean ergodic theorem to the problem of Markoff's process. *Proc. Imp. Acad. Japan* 14 (333-339).

[51] YOSIDA, K. (1938/39). Operator-theoretical treatment of the Markoff's process. *Proc. Imp. Acad. Japan* 14 (1938) (363-367) and 15 (1939) (127-130).

[52] YOSIDA, K. *and* KAKUTANI, S. (1941). Operator-theoretical treatment of Markoff's process and mean ergodic theorem. *Ann. of Math. (2)* 42 (188-228).

[53] YUKICH, J. E. (1985). Sufficient conditions for the uniform convergence of means to their expectations. *Sankhya*, Ser. A 47 (203-208).

Goran Peskir
Department of Mathematical Sciences
University of Aarhus, Denmark
Ny Munkegade, DK-8000 Aarhus
home.imf.au.dk/goran
goran@imf.au.dk

Michel Weber
I.R.M.A. Unité de Recherche
associée C.N.R.S., 1
7, rue René Descartes
67084 Strasbourg
France