Ergod. Th. & Dynam. Sys. (2002), **22**, 1001–1015
DOI: 10.1017/S0143385702000731
© *2002 Cambridge University Press*
Printed in the United Kingdom

Ergodicity and mixing via Young measures

ZVI ARTSTEIN† and MICHAEL GRINFELD‡

† *Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel*
(e-mail: zvi.artstein@weizmann.ac.il)

‡ *Department of Mathematics, University of Strathclyde, Glasgow G1 1XH, UK*
(e-mail: michael@maths.strath.ac.uk)

(*Received* 3 *October* 2001 *and accepted in revised form* 20 *December* 2001)

Abstract. Connections are established between mixing or ergodic properties of maps on the one hand, and the convergence of the iterates of the map, or of the empirical measures of the iterates, to a constant measure-valued map, on the other. The uniqueness of an absolutely continuous ergodic measure can also be verified via the convergence. The technique helps to identify ergodic and mixing pairs and verify the uniqueness in specific examples.

1. *Introduction*

This paper examines the relations between mixing or ergodicity properties of a map and the convergence of the iterates of the maps or of the empirical measures of the iterates. The convergence that we use is to probability measure-valued maps. Such maps are referred to in the literature by various terms, e.g. Markov transitions maps, stochastic kernels and Young measures. We adopt the last terminology since it relates to the dynamical systems characteristics of the process.

Young measures serve as appropriate limits in a completion of function spaces when instantaneous oscillations occur in limits of solutions and parameters of variational problems and dynamical systems. These measures were introduced to the calculus of variations by Young, and to control theory by Warga (see Young [**18**], Warga [**17**] and references therein). Further applications of Young measures as limits of point-valued maps were worked out in the framework of partial differential equations and singularly perturbed ordinary differential equations (see Tartar [**14**], Ball [**5**], Artstein and Vigodner [**3**], Artstein [**1**], and Pedregal [**13**]).

The connection between Young measures and invariance and ergodic properties worked out in this paper concern dynamical systems generated by a map from a space to itself. We consider Young measures associated with the iterates of the map and with the empirical measures of the iterates of the map. The convergence of the latter to a constant-valued Young measure is associated with ergodicity, while the convergence of the iterates themselves is associated with mixing. The convergence of the empirical measures in the

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

case of ergodicity is practically a known result, and holds without any particular regularity conditions. The convergence of the iterates in the case of mixing was pointed out by Grinfeld and Knight [**10**], under some restrictive conditions. We extend this result here and show that the convergence holds in general. The main results of the paper, however, establish the converse direction for both mixing and ergodicity. This is carried out under an additional continuity condition, namely that the measure of the discontinuity set is zero. We show that this additional condition cannot be removed.

The characterizations of ergodicity and of mixing enable us to identify ergodic and mixing pairs in the case that the iterates converge to a constant-valued Young measure and the additional continuity assumption is satisfied. Then the limit value is ergodic or mixing when it is not orthogonal to the underlying measure. We offer examples demonstrating that the non-orthogonality condition cannot be removed.

The paper is organized as follows. In §2 we display our setting and terminology, recall the notion of Young measures and some of their properties, and recall the continuity property, namely μ-continuity, which plays a crucial role in our analysis. In §3 we examine the connection between the Young measures and ergodicity. In §4 we establish the connection between Young measures and mixing. In both, §§3 and 4, we provide counter-examples demonstrating that the respective continuity and non-orthogonality assumptions cannot be removed. The closing section is devoted to further examples demonstrating the applicability of the technique to some well-known maps.

2. *The setting, Young measures and μ-continuity*

This section covers three preliminary subjects. First, we display the setting and terminology used throughout the paper. Then we recall the notion of, and list some basic facts about, Young measures. At the end of this section we recall the notion of μ-continuity of a map and draw a consequence.

Assumption 2.1. Throughout the paper X is a separable metric space endowed with its Borel σ-field.

We consider the measurable space X with various probability measures and employ the theory of spaces of probability measures on metric spaces. Our references are Bertsekas and Shreve [**6**, Ch. 7] and Billingsley [**7**, Ch. 1, §2].

Consider (X, λ) where λ is a probability measure. Let Y be a separable metric space and let $T : X \to Y$ be measurable. The measure induced from (X, λ) on Y by T is the probability measure $T\lambda$ defined by

$$(T\lambda)(A) = \lambda(T^{-1}(A)). \tag{2.1}$$

The measure induced by T is also referred to as the distribution of T with respect to λ. Now let Z be a separable metric space, and let $S : Y \to Z$ be measurable. It is clear that $(ST)\lambda = S(T\lambda)$.

Denote by $\mathcal{P}(X)$ the space of probability measures on the space X. We consider the space $\mathcal{P}(X)$ with the weak convergence of measures. The weak convergence on the space $\mathcal{P}(X)$ can be represented as a separable metric space (see [**6**, Proposition 7.20]).

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

The convergence is determined as follows. A sequence μ_i converges to μ_0 if

$$\int_X \gamma(x)\mu_i(dx) \to \int_X \gamma(x)\mu_0(dx) \tag{2.2}$$

for every bounded and continuous function $\gamma : X \to R$, where R is the real line.

Let $T : X \to X$ be measurable. A probability measure μ on X is invariant with respect to T, or T is preserving μ, if $T\mu = \mu$.

We need to consider probability measures on product spaces, say on $X \times Y$. We use boldface letters to denote such measures. Let $\boldsymbol{\mu}$ be such a measure. The marginal of $\boldsymbol{\mu}$ on Y is the probability measure, say ν, defined by

$$\nu(C) = \boldsymbol{\mu}(X \times C), \tag{2.3}$$

and likewise with the marginal on X. It is clear (e.g. from the criterion (2.2)) that the marginal operation is a continuous mapping from $\mathcal{P}(X \times Y)$ to $\mathcal{P}(Y)$.

We turn now to Young measures. The literature offers various representations of Young measures; we elaborate here on the one we find suitable for our needs.

Let X and Y be separable metric spaces, and let λ be a probability measure on X. A Young measure from (X, λ) to Y is a measurable map

$$\mu(\cdot) : X \to \mathcal{P}(Y), \tag{2.4}$$

where the metric of weak convergence is taken on the image space $\mathcal{P}(Y)$. The measurability of $\mu(\cdot)$ is characterized by the measurability of each of the mappings $x \to \mu(x)(C)$ for every measurable subset C of Y (see, e.g., Bertsekas and Shreve [**6**, Proposition 7.26]). The functions $f : X \to Y$ are regarded as (degenerate) Young measures by associating with f the Young measure whose value at x is the Dirac measure $\delta_{f(x)}$; namely, the measure supported on the point $\{f(x)\}$. We use the notation $f(\cdot)$ or the boldface font $\mathbf{f}$ when we wish to examine f as a Young measure.

The family of Young measures from (X, λ) to Y is denoted by $\mathcal{Y}((X, \lambda), Y)$.

We identify an element $\mu(\cdot)$ in $\mathcal{Y}$ with its direct integral, namely the probability measure $\boldsymbol{\mu}$ on $X \times Y$ determined by the relation

$$\boldsymbol{\mu}(B \times C) = \int_B \mu(C)\lambda(dx) \tag{2.5}$$

for all measurable subsets $B \subseteq X$ and $C \subseteq Y$. We alternate freely between the two representations of a Young measure, namely between $\mu(\cdot)$ and $\boldsymbol{\mu}$.

The space of Young measures $\mathcal{Y}((X, \lambda), Y)$ thus becomes a subspace of $\mathcal{P}(X \times Y)$. In particular, a criterion for convergence of $\boldsymbol{\mu}_i$ to $\boldsymbol{\mu}_0$ in $\mathcal{Y}((X, \lambda), Y)$ is similar to (2.2), namely, that

$$\int_{X\times Y} \gamma(x, y)\boldsymbol{\mu}_i(dx \times dy) \to \int_{X\times Y} \gamma(x, y)\boldsymbol{\mu}_0(dx \times dy) \tag{2.6}$$

whenever $\gamma(x, y) : X \times Y \to R$ is bounded and continuous. Using, however, the representation of Young measures as maps, we get the condition for convergence stated as follows:

$$\int_X \int_Y \gamma(x, y)\mu_i(x)(dy)\lambda(dx) \to \int_X \int_Y \gamma(x, y)\mu_0(x)(dy)\lambda(dx). \tag{2.7}$$

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

Given this representation, it is easy to see that for Young measures in $\mathcal{Y}((X,\lambda),Y)$, the weak convergence can be characterized by the convergence (2.7) for functions $\gamma(x,y)$ which are bounded, continuous in y and may be merely measurable in the x variable. Expression (2.7) also reflects the role of the base measure λ in the convergence, as λ stays put in the course of the convergence.

The previous characterization of convergence of Young measures implies the following observation which will become handy in further developments.

LEMMA 2.2. *Let $\mu_i(\cdot)$ be Young measures which converge in $\mathcal{Y}((X,\lambda),Y)$ to $\mu_0(\cdot)$. Let ν be a probability measure on X, absolutely continuous with respect to λ. Then the convergence of $\mu_i(\cdot)$ to $\mu_0(\cdot)$ holds also when the Young measures are interpreted as elements of $\mathcal{Y}((X,\nu),Y)$.*

Proof. Verifying (2.7) with ν replacing λ, is equivalent to verifying (2.7) with λ when $\gamma(x,y)$ is multiplied by $d\nu/d\lambda(x)$. The existence of the latter is guaranteed by the absolute continuity. Since the kernel in (2.7) is allowed to be merely measurable in the x variable, the claim follows. □

It is clear that the absolute continuity condition in the previous lemma cannot be dropped.

In analogy with the case of mappings, we define the measure induced from (X,λ) on Y by the Young measure $\mu(\cdot)$ to be the probability measure $\boldsymbol{\mu}\lambda$ defined by

$$(\boldsymbol{\mu}\lambda)(A) = \int_X \mu(x)(A)\lambda(dx). \tag{2.8}$$

This expression coincides with (2.1) when the Young measure is degenerate, namely its values are Dirac measures. (Note that we abuse formality, as $\mu(\cdot)$ is actually a mapping into $\mathcal{P}(Y)$, hence formally the induced measure should be in $\mathcal{P}(\mathcal{P}(Y))$.) The induced measure will also be called the distribution of the Young measure. The Young measure $\boldsymbol{\mu}$ is called measure preserving if $\boldsymbol{\mu}\lambda = \lambda$. A comprehensive examination of measure-preserving Young measures (called polymorphisms) was carried out by Vershik [**16**].

We need the following observation for further reference.

Remark 2.3. The distribution of a Young measure coincides with the marginal of the Young measure on the image space. Likewise, the distribution of a mapping $T : X \to Y$ with respect to a probability measure λ on X, namely $T\lambda$, coincides with the marginal on Y of the Young measure $\mathbf{T} = T(\cdot)$.

We close this section on preliminaries with the definition of μ-continuity. Consider the following notation. Let X and Y be metric spaces and let $g : X \to Y$ be measurable. The set of points at which g is discontinuous is denoted by $\mathrm{Disc}(g)$.

Definition 2.4. Let X and Y be separable metric spaces and let $T : X \to Y$ be measurable. Let μ be a probability measure on X. We say that T is μ-continuous if $\mu(\mathrm{Disc}(T)) = 0$. Clearly, a continuous map is μ-continuous for any measure μ.

The μ-continuity notion (under different titles) has been found to be useful in various circumstances (see, for instance, Billingsley [**7**, §5] and Topsøe [**15**] for the theory, and Artstein and Wets [**4**] for an application). The following property is pivotal for us.

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

LEMMA 2.5. *Let $T : X \to Y$ be measurable, with X and Y separable metric spaces. The induced measure operation $T : \mathcal{P}(X) \to \mathcal{P}(Y)$ is continuous with respect to the weak convergence of measures at each element μ such that T is μ-continuous.*

Proof. See Billingsley [**7**, Theorem 5.1]. □

COROLLARY 2.6. *Let $T : X \to X$ be measurable. Let λ be a probability measure on X. Suppose that $T^n\lambda$ converge as $n \to \infty$, say to μ, and suppose that T is μ-continuous. Then μ is an invariant measure with respect to T.*

Proof. Both $T^n\lambda$ and $T^{n+1}\lambda$ converge to μ, and $TT^n\lambda = T^{n+1}\lambda$. The continuity established in the previous lemma implies that $T\mu = \mu$. □

3. *Young measures and ergodicity*

In this section we associate the ergodic property of dynamical systems with the Young measure limit of empirical measures of iterates. First, we recall the relevant definitions.

Let μ be a probability measure on X and let $T : X \to X$ be measurable and μ-preserving. The pair (T, μ) is called ergodic if $T^{-1}(B) = B$ implies that either $\mu(B) = 0$ or $\mu(B) = 1$ (see, e.g., Mãné [**11**]). (Note that the measure-preserving property is part of the definition; non-existence of invariant sets may hold for non-measure-preserving maps.)

Let $T : X \to X$ be measurable. We denote by $\sigma_T^n(x)$ the empirical measure induced by the first $n-1$ iterates of T applied to the point x, namely the probability measure equally distributed on the finite set $\{x, T(x), T^2(x), \ldots, T^{n-1}(x)\}$. In particular, we consider $\sigma_T^n(\cdot)$ as a Young measure, and, as usual, we may also use the boldface notation and denote it by $\boldsymbol{\sigma}_T^n$.

The following invariance property plays a role in our analysis. Example 3.7 below shows that the μ-continuity cannot be removed from the statement of the lemma.

LEMMA 3.1. *Let $T : X \to X$ be measurable. Let λ be a probability measure on X. Suppose that $\sigma_T^n(\cdot)$ converge in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\mu}$, with a constant value, say $\mu(x) = \mu$ for λ-almost every x, and suppose that T is μ-continuous. Then μ is an invariant measure of T.*

Proof. It is enough to prove that

$$\int_X \gamma(x)\mu(dx) = \int_X \gamma(x)(T\mu)(dx) \tag{3.1}$$

for every $\gamma : X \to X$ bounded and continuous. By the definition of induced measure (see (2.1)), the map T on the right-hand side of (3.1) can be shifted into the x-argument. Hence, it is enough to verify that

$$\int_X \gamma(x)\mu(dx) = \int_X \gamma(T(x))\mu(dx) \tag{3.2}$$

for every bounded and continuous γ. The convergence of $\sigma_T^n(\cdot)$ to $\mu(\cdot) \equiv \mu$ implies that the left-hand side of (3.2) is equal to

$$\lim_{n\to\infty} \int_X \frac{1}{n} \sum_{j=0}^{n-1} \gamma(T^j(x))\lambda(dx). \tag{3.3}$$

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

Since T is μ-continuous and γ is continuous, it follows that γT is μ-continuous, and hence (by Lemma 2.5) the right-hand side of (3.2) is equal to

$$\lim_{n\to\infty} \int_X \frac{1}{n} \sum_{j=0}^{n-1} \gamma(T^{j+1}(x))\lambda(dx). \tag{3.4}$$

The difference between the values of (3.3) and (3.4) is of order n^{-1}. This completes the proof. □

The following observation will be useful in the sequel.

LEMMA 3.2. *Let μ be a probability measure on X, and let C be a measurable subset of X. For every $\varepsilon > 0$ there is a closed set $K_\varepsilon \subseteq C$ and there is an open set $Q_\varepsilon \supseteq C$, such that $\mu(Q_\varepsilon \backslash K_\varepsilon) < \varepsilon$, and moreover, $\mu(\partial Q_\varepsilon) = 0$, where ∂ denotes the boundary of the set.*

Proof. Without the condition $\mu(\partial Q_\varepsilon) = 0$ the result is verified in [**7**, Ch. 1, Theorem 1.1]. Taking then, say, the open η-neighborhood of Q_ε for small η will not change the estimates, and among the continuum of such η one can be found which satisfies the boundary condition. □

Here is the first main result of this section. Example 3.7 and Remark 3.9 below show that the μ-continuity cannot be removed from the statement of the theorem.

THEOREM 3.3. *Let μ be a probability measure on X. Let $T : X \to X$ be measurable. A necessary condition for the pair (T, μ) to be ergodic is the convergence of the empirical measures $\sigma_T^n(\cdot)$ as Young measures in $\mathcal{Y}((X, \mu), X)$, to the Young measure, say $\boldsymbol{\mu}$, with a constant value equal to μ. The convergence is also a sufficient condition in the case where T is μ-continuous.*

Proof. Sufficiency. We show that (T, μ) is ergodic by showing that T is μ-measure preserving, and that there does not exist a subset B which is invariant with respect to T, yet $0 < \mu(B) < 1$. Since in the sufficiency case we assume that T is μ-continuous, the invariance of μ follows from Lemma 3.1. If there exists a set B as described, then for $x \in B$, and for all n, the support of $\sigma_T^n(x)$ is in B. Furthermore, let ε be small, and let K be a closed subset of B such that $\mu(B\backslash K) < \varepsilon$, as guaranteed in Lemma 3.2. Since T is μ-measure preserving, it follows that $\sigma_T^n(\cdot)$ is also μ-measure preserving (for the definition of the induced measure, see §2), hence $\boldsymbol{\sigma}_T^n(K \times K) \geq \mu(B) - \varepsilon$. By, for example, [**7**, Ch. 1, Theorem 2.1(iii)] if the convergence of $\boldsymbol{\sigma}_T^n$ to $\boldsymbol{\mu}$ holds, then $\boldsymbol{\mu}(K \times K) \geq \mu(B) - \varepsilon$. For ε small, the latter contradicts (2.5), which implies in the case where $\mu(x) \equiv \mu$ that $\boldsymbol{\mu}(B \times B) = \mu(B)\mu(B)$.

Necessity. Note that ergodicity implies (see, e.g., [**11**, Ch. II, Proposition 2.1]) that when applying criterion (2.7) to check the convergence of $\sigma_T^n(\cdot)$ to $\boldsymbol{\mu}$, the outer integrand converges pointwise (even for functions merely integrable in the y-variable). Hence (2.7) verifies the convergence of $\boldsymbol{\sigma}_T^n$ to $\boldsymbol{\mu}$. This completes the proof. □

The characterization in the previous theorem assumes that X is already endowed with a probability measure with respect to which T is ergodic. In many applications, however, one starts with a given underlying probability and looks for ergodic pairs. The second main

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

result of this section relates such ergodic measures to iterates. Recall that the probability measures μ and λ are orthogonal if there exists a set B such that $\mu(B) = 1$ while $\lambda(B) = 0$. Example 3.8 below shows that the non-orthogonality condition cannot be dropped from the statement of the result.

THEOREM 3.4. *Let (X, λ) be a probability measure space. Let $T : X \to X$ be measurable. Suppose that $\sigma_T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\mu}$, with a constant value, say $\mu(x) = \mu$ for λ-almost every x. If μ is not orthogonal to λ and T is μ-continuous, then (T, μ) is an ergodic pair.*

Proof. By Lemma 3.1, T is μ-measure preserving. We have therefore to check that no subset B can be found which is invariant under T and such that $0 < \mu(B) < 1$. Suppose that such a set B exists. Let μ_1 be the absolutely continuous part of μ (with respect to λ). Then either $\mu_1(B) \neq 0$ or $\mu_1(X \backslash B) \neq 0$. Suppose that $\mu_1(B) \neq 0$. Denote the restriction of μ_1 to B by μ_2. This restriction is absolutely continuous with respect to λ. By Lemma 2.2, $\sigma_T^n(\cdot)$ converge also in $\mathcal{Y}((X, \mu_2), X)$ to $\mu(x) \equiv \mu$. For ε small let Q_ε be the open set guaranteed by Lemma 3.2. In particular, $B \subseteq Q_\varepsilon$ and $\mu(Q_\varepsilon) < 1$. The invariance of B with respect to T implies that for μ_2-almost every x the support of $\sigma_T^n(x)$ is in $B \times Q_\varepsilon$. The convergence to the constant Young measure $\mu(\cdot)$ implies (by [**7**, Ch. 1, Theorem 2.1(v)]) that for μ_2-almost every x we have $\mu(x)(Q_\varepsilon) = 1$. This contradicts the construction of Q_ε and completes the proof. □

COROLLARY 3.5. *Let (X, λ) be a probability measure space and let $T : X \to X$ be measurable. Suppose that $\sigma_T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\nu}$, with a constant value, say $\nu(x) = \nu$ for λ-almost every x. Then any ergodic pair (T, μ) with μ not orthogonal to λ satisfies $\mu = \nu$. In particular, there exists at most one such ergodic pair.*

Proof. Let (T, μ) be an ergodic pair as described, and let μ_0 be the absolutely continuous part of μ with respect to λ. By Lemma 2.2 the empirical measures $\sigma_T^n(\cdot)$ converge to $\boldsymbol{\nu}$ also in $\mathcal{Y}((X, \mu_0), X)$. In view of Theorem 3.3, the empirical measures $\sigma_T^n(\cdot)$ converge in $\mathcal{Y}((X, \mu), X)$ to the Young measure with constant value μ. Since μ_0 is absolutely continuous also with respect to μ, it follows that the convergence also holds in $\mathcal{Y}((X, \mu_0), X)$. Hence $\mu = \nu$ as claimed. □

In Theorem 3.4 and its corollary, the convergence of the empirical measures of the iterates to a constant valued Young measure was assumed. We can get similar consequences under a weaker assumption, as follows.

PROPOSITION 3.6. *Let (X, λ) be a probability measure space and let $T : X \to X$ be measurable. Suppose that $\sigma_T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\nu}$, which is constant-valued on a subset B of X, say $\nu(x) = \nu$ for $x \in B$. Suppose also that $\lambda(B) > 0$, and suppose that ν is not orthogonal to the restriction of λ to B. If T is ν-continuous then the pair (T, ν) is ergodic. Furthermore, (T, ν) is the only ergodic pair with ν not orthogonal to the restriction of λ to B.*

Proof. Apply Theorem 3.4 and Corollary 3.5 to the restriction of λ to B. □

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

Now we display the counter-examples promised above.

Example 3.7. Let $X = [0, 1]$, and define $T(x) = \frac{1}{2}x$ when $x \neq 0$ and $T(0) = 1$. Then for every x the iterates $T^n(x)$ converge as $n \to \infty$ to 0. Hence the empirical measures converge as Young measures to the constant-valued Young measure supported on $\{0\}$, regardless of the probability measure X is endowed with. In particular, when X is endowed with the Dirac measure δ_0 supported on $\{0\}$, the measure is not an invariant measure of T and (T, δ_0) is not an ergodic pair. This is due to the discontinuity of T (see also Remark 3.9).

Example 3.8. Consider $X = [0, 1] \times S^1$ with S^1 being the unit circle. We parameterize S^1 as the unit interval, identifying 1 with 0, and consider the summation operation modulus 1. We endow X with the Lebesgue measure. We define now the mapping $T : X \to X$ as follows. Consider the decreasing sequence $\alpha_i = 2^{-i}$ for $i = 1, 2, \ldots$ and partition each interval $[\alpha_{i+1}, \alpha_i]$ into subintervals of equal length, say length h_i, such that $h_i^{-1}\alpha_i$ tends to ∞, say $h_i^{-1}\alpha_i = 2^i$. For (ξ, θ) with $\alpha_{i+1} + kh_i \leq \xi \leq \alpha_{i+1} + (k + 1)h_i$ define $T(\xi, \theta) = (\xi - h_i, \theta + \alpha_i)$, this for $k = 1, \ldots, 2^i - 1$. This defines T except when $\alpha_{i+1} < \xi < \alpha_{i+1} + h_i$, and when $\xi = 0$. In the first case we extend T linearly between its values at $\xi = \alpha_{i+1}$ and $\xi = \alpha_{i+1} + h_i$. For $\xi = 0$ we define $T(0, \theta) = (0, \theta)$.

The map T defined above is continuous. The choice of h_i, specifically the choice $h_i\alpha_i = 2^i$, with the fact that the θ-coordinate is shifted by the amount α_i when $\alpha_{i+1} + h_i < \xi \leq \alpha_i$, imply that for every $\xi > 0$ the iterates $T^n(x)$ for $n = 1, \ldots, N$ with large N, are spread close to evenly around the circle $\{0\} \times S^1$. Hence the empirical measures of the iterates converge to a constant valued Young measure which is the Lebesgue measure on $\{0\} \times S^1$. The convergence of $\sigma_T^n(x)$ holds almost everywhere, yet T is not ergodic with respect to the limit measure. Indeed, the subset $\{0\} \times S^1$ consists of fixed points of T. (This does not contradict Theorem 3.4 since the Lebesgue measure on the circle $\{0\} \times S^1$ is orthogonal to the Lebesgue measure on X.)

Remark 3.9. The cause for the irregularity in Example 3.7 is the discontinuity of T at zero. One may attempt to correct the discontinuity by either redefining $T(0)$ or closing the graph. In the latter case one may employ notions for invariance for set-valued maps, as defined, e.g., in Miller and Akin [**12**] or in Artstein [**2**]. In Example 3.7 such considerations would resolve the invariance issue, and produce an ergodic pair, namely the one generated by the fixed point $T(0) = 0$. A variant of Example 3.8 would show that ergodicity cannot be recovered by closing the graph. Indeed, if we change the definition of T in Example 3.8 for points $(0, \theta)$ and have $T((0, \theta)) = (1, \theta)$, we would get that the convergence of the iterates to the Lebesgue measure on the circle holds for every $x \in X$. Hence, the convergence would hold everywhere even if the circle $\{0\} \times S^1$ is assigned with a positive measure. Yet, invariance does not hold, due to the lack of continuity. Redefining T or closing its graph would lead back to Example 3.8, where invariance holds but ergodicity fails.

4. *Young measures and mixing*

In this section we associate the mixing property of a dynamical system with the Young measure limit of iterates.

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

First we recall the definition of mixing (see, e.g., Mãné [**11**]). Let (X, μ) be a probability measure space. Let $T : X \to X$ be measurable. Recall that (T, μ) is called mixing if

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) \to \mu(A)\mu(B) \tag{4.1}$$

for every pair of Borel sets A and B in X. If (T, μ) is mixing then μ is an invariant measure of T.

We first establish the invariance of the Young measure limit of iterates. Example 3.7 shows that the μ-continuity cannot be removed from the statement of the lemma.

LEMMA 4.1. *Let $T : X \to X$ be measurable. Let λ be a probability measure on X. Suppose that as Young measures, the iterates $T^n(\cdot)$ converge in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\mu}$, with a constant value, say $\mu(x) = \mu$ for λ-almost every x. If T is μ-continuous, then μ is an invariant measure of T.*

Proof. Convergence of the iterates $T^n(\cdot)$ implies in particular the convergence of the empirical measures $\sigma_T^n(\cdot)$ to the same limit. Hence the result follows from Lemma 3.1. A direct proof is even easier. The measure induced from (X, λ) by T^n on X is the marginal of $T^n(\cdot)$ when interpreted as a Young measure, see Remark 2.3. Since the marginal of $\boldsymbol{\mu}$ is the constant value μ, and since both $T^n\lambda$ and $T^{n+1}\lambda$ converge to μ, the invariance follows from the continuity established in Corollary 2.6. □

Here is the first main result of this section. Example 3.7 of the previous section shows that here also, the μ-continuity cannot be removed from the statement of the result.

THEOREM 4.2. *Let μ be a probability measure on X. Let $T : X \to X$ be measurable. A necessary condition for the pair (T, μ) to be mixing is the convergence of the iterates $T^n(\cdot)$ as Young measures in $\mathcal{Y}((X, \mu), X)$, to the Young measure with a constant value equal to μ. The convergence is also a sufficient condition in case T is μ-continuous.*

Proof. Sufficiency. We assume that T is μ-continuous and $T^n(\cdot)$ converge to $\boldsymbol{\mu}$ with $\mu(x) = \mu$ for μ-almost every x. We have to verify (4.1).

We claim that it is enough to establish (4.1) when B is an open set with $\mu(\partial B) = 0$. Indeed, let C be a general Borel set. For every $\varepsilon > 0$, let Q_ε be the open set guaranteed by Lemma 3.2. Since, in particular, $C \subseteq Q_\varepsilon$, it follows that $\mu(T^{-n}(A) \cap Q_\varepsilon) - \mu(T^{-n}(A) \cap C) < \varepsilon$. If (4.1) is true whenever B in (4.1) is one of the open sets Q_ε with an arbitrarily small ε, it follows that it holds in the limit, namely for the Borel set C. Since C is an arbitrary Borel set, the claim follows.

Next, we claim that it is enough to establish (4.1) for open sets A with $\mu(\partial A) = 0$. Indeed, given a Borel set C, for every $\varepsilon > 0$ let Q_ε be the open set guaranteed in Lemma 3.2. Then, $T^{-n}(Q_\varepsilon) \cap B = (T^{-n}(C) \cap B) \cup (T^{-n}(Q_\varepsilon \backslash C) \cap B)$, and the union in the formula is disjoint. At this point we use the μ-continuity of T, specifically the invariance established in Lemma 4.1, and deduce from $\mu(Q_\varepsilon \backslash C) < \varepsilon$ that $\mu(T^{-n}(Q_\varepsilon \backslash C)) < \varepsilon$. Hence, in particular, $\mu(T^{-n}(Q_\varepsilon \backslash C) \cap B) < \varepsilon$, which implies that $\mu(T^{-n}(Q_\varepsilon)) - \mu(T^{-n}(C)) < \varepsilon$. If (4.1) holds for all such Q_ε, it holds in the limit as $\varepsilon \to 0$, and hence it holds for the general Borel set C, namely for all Borel sets.

In view of the preceding reductions, it is enough to establish (4.1) when A and B are open sets in X with boundaries of μ-measure zero. For such sets the product $B \times A$ is open

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

in $X \times X$ and its boundary is the union of $\partial B \times A$ and $B \times \partial A$. Applying (2.5) to the Young measure $\boldsymbol{\mu}$, and bearing in mind that $\boldsymbol{\mu} = \mu(\cdot)$ is constant-valued with a value equal to μ, it follows that $\boldsymbol{\mu}(\partial(B \times A)) = 0$. Let $\mathbf{T}^n$ be the Young measure associated with T^n. Since $\mathbf{T}^n$ converge to $\boldsymbol{\mu}$ it follows (see [**7**, Ch. 1, Theorem 2.1(v)]) that $\mathbf{T}^n(B \times A)$ converge to $\boldsymbol{\mu}(B \times A)$. Since $\boldsymbol{\mu}$ is constant valued and its value μ is the underlying probability measure on X, it follows that $\boldsymbol{\mu}(B \times A)$ is equal to $\mu(A)\mu(B)$. The value $\mathbf{T}^n(B \times A)$ is none other than the term $\mu(T^{-n}(A) \cap B)$ appearing on the left-hand side of (4.1). This verifies the sufficiency of the condition.

Necessity. We assume the mixing. It is clear that a Young measure $\boldsymbol{\mu} = \mu(\cdot)$ in $\mathcal{Y}((X, \lambda), Y)$ assumes a constant value, say ν, if and only if $\boldsymbol{\mu}(B \times A) = \nu(A)\lambda(B)$ for all Borel sets B and A in X and Y, respectively. This claim follows directly from (2.5). It is also clear that for the 'if' part it is enough that the equality $\boldsymbol{\mu}(B \times A) = \nu(A)\lambda(B)$ holds for sets A and B which are open, and such that $\lambda(\partial B) = 0$ and $\nu(\partial A) = 0$. Indeed, given a general product $B \times A$, the set B and A can be approximated, respectively, in the λ and ν measures by open sets with boundaries of zero measure (as guaranteed in Lemma 3.2). Passing then to the limit establishes the equality for $B \times A$. To check the necessity part of the theorem, we have to take $X = Y$ and $\lambda = \nu = \mu$ and check that $\boldsymbol{\mu}(B \times A) = \mu(A)\mu(B)$ for open sets with a boundary of μ-measure zero. This holds, as for such sets $\boldsymbol{\mu}(B \times A)$ is the limit of $\mu(T^{-n}(A) \cap B)$ appearing on the left-hand side of (4.1), and the mixing property implies that the convergence to $\mu(A)\mu(B)$ holds. This completes the proof. □

Before turning to the second main result of the section, we establish a useful connection between Young measures, ergodicity and mixing. The result reflects the property that an ergodic map which exhibits characteristics of mixing on part of the space, is in fact mixing.

PROPOSITION 4.3. *Let μ be a probability measure on X and let $T : X \to X$ be measurable. Suppose that (T, μ) is ergodic. Suppose further that on a Borel subset K with positive μ-measure, the iterates $T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((K, \mu), X)$ to a constant-valued measure, say $\nu(\cdot) \equiv \nu$ on K, such that T is ν-continuous. Then $\nu = \mu$ and (T, μ) is mixing.*

Proof. Let $K_1 = T^{-1}(K)$. Since T is ergodic with respect to μ and, in particular, the measure μ is invariant, it follows that the restriction of μ to K is the measure induced by T from the restriction of μ to K_1. From the definition of the Young measure limit of iterates, it follows that the iterates $T^n(\cdot)$ converge also in $\mathcal{Y}((K_1, \mu), X)$ to the same constant-valued Young measure. Consequently, if we denote $C_1 = K \cup K_1$, then the convergence of the iterates holds in $\mathcal{Y}((C_1, \mu), X)$. Inductively, let $K_{j+1} = T^{-1}(C_j)$, for $j = 1, 2, \ldots$; then the iterates $T^n(\cdot)$ converge in $\mathcal{Y}((C_{j+1}, \mu), X)$, as $n \to \infty$, to $\nu(\cdot) \equiv \nu$. By ergodicity the union of the increasing sequence C_j exhausts the measure μ. Hence the iterates $T^n(\cdot)$ converge in $\mathcal{Y}((X, \mu), X)$ to the constant ν. By Theorem 3.3, the empirical measures of the iterates converge in $\mathcal{Y}((X, \mu), X)$ to the constant μ. Since convergence of the iterates clearly implies the convergence of the empirical measures, and to the same value, it follows that $\nu = \mu$, and by Theorem 4.2 the mixing property is guaranteed. □

The characterization in Theorem 4.2 assumes that X is already endowed with the probability measure with respect to which T is mixing. As was done in the previous

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

section in the case of ergodicity, we are interested in the situation where one starts with a given underlying probability and looks for mixing pairs. The following result, which is the second main result of the section, relates such mixing pairs to iterates. Orthogonality of measures was recalled in the previous section. Example 4.7 below shows that the non-orthogonality condition cannot be dropped from the statement of the result.

THEOREM 4.4. *Let (X, λ) be a probability measure space. Let $T : X \to X$ be measurable. Suppose that $T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\mu}$, with a constant value, say $\mu(x) = \mu$ for λ-almost every x. If μ is not orthogonal to λ and T is μ-continuous, then (T, μ) is mixing.*

Proof. The convergence of $T^n(\cdot)$ to $\boldsymbol{\mu}$ implies the convergence of the empirical measures $\sigma_T^n(\cdot)$ to the same Young measure limit. In view of Theorem 3.4 the pair (T, μ) is ergodic. Let μ_0 be the absolutely continuous part of μ (with respect to λ). The assumption is that $\mu_0(X) > 0$. Let K be the Borel subset of X such that $\mu_0(K) = \mu(K)$ and $\lambda(X \backslash K) = 0$. The existence of such a set K follows from the Hahn–Jordan decomposition. In particular, $\mu_0 = \mu$ on K. By Lemma 2.2 the iterates $T^n(\cdot)$ converge in $\mathcal{Y}((K, \mu_0), X)$ to the constant-valued Young measure $\mu(x) = \mu$. By Proposition 4.3 the ergodic pair (T, μ) is actually mixing. This completes the proof. □

The following corollary of our result (see also Proposition 5.5 below) agrees with, and in fact partially generalizes, many results and observations concerning the uniqueness of an invariant measure of a chaotic map which is absolutely continuous with respect to a physical measure (see, e.g., Mañé [**11**, Ch. III, Theorem 1.2], Bowen [**9**, Theorems 4.1 and 4.14]).

COROLLARY 4.5. *Let (X, λ) be a probability measure space and let $T : X \to X$ be measurable. Suppose that $T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\nu}$, with a constant value, say $\nu(x) = \nu$, for λ-almost every x. If μ is a probability measure on X such that (T, μ) is mixing on X and μ is not orthogonal to λ, then $\mu = \nu$. In particular, the convergence of $T^n(\cdot)$ to a constant-valued Young measure implies that there exists, at most, one probability measure μ on X which is not orthogonal to λ and such that (T, μ) is mixing.*

Proof. Let μ_0 be the absolutely continuous part of μ with respect to λ. By Lemma 2.2 the convergence of the iterates $T^n(\cdot)$ to $\boldsymbol{\nu}$ in $\mathcal{Y}((X, \lambda), X)$, assumed in the claim, holds also in $\mathcal{Y}((X, \mu_0), X)$. By Theorem 4.2, these iterates $T^n(\cdot)$ converge in $\mathcal{Y}((X, \mu), X)$ to the Young measure with constant value μ. Since μ_0 is also absolutely continuous with respect to μ, it follows again from Lemma 2.2 that these iterates converge to $\boldsymbol{\nu}$ in $\mathcal{Y}((X, \mu_0), X)$ as well. In particular, $\mu = \nu$ as claimed. □

In Theorem 4.4 and its corollary, the convergence of the iterates to a constant valued Young measure was assumed. As was done in Proposition 3.6 concerning ergodicity, here as well one can get similar consequences under a weaker assumption, as follows.

PROPOSITION 4.6. *Let (X, λ) be a probability measure space and let $T : X \to X$ be measurable. Suppose that $T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\nu}$, which is constant-valued on a subset, say B, of X, say $\nu(x) = \nu$ for $x \in B$.*

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

Suppose also that $\lambda(B) > 0$, and that ν is not orthogonal to the restriction of λ to B. If T is ν-continuous then the pair (T, ν) is mixing. Furthermore, (T, ν) is the only mixing pair with T being ν-continuous and ν not orthogonal to the restriction of λ to B.

Proof. Apply Theorem 4.4 and Corollary 4.5 to the restriction of λ to B. □

Example 4.7. Employing a variant of Example 3.8 we show here that the non-orthogonality condition cannot be removed from Theorem 4.4. Consider $X = [0, 1] \times S^1$ parameterized as in Example 3.8, and employ the same partition system as in that example. This time define T by $T(\xi, \theta) = (\rho(\xi), \eta(\xi, \theta))$ given as follows. For (ξ, θ) with $\alpha_{i+1} + kh_i \le \xi \le \alpha_{i+1} + (k+1)h_i$ we define $\rho(\xi) = \xi - h_i$, this for $k = 1, \ldots, 2^i - 1$. This defines $\rho(\xi)$ except when $\alpha_i < \xi < \alpha_i + h_i$, and when $\xi = 0$. In the first case we extend $\rho(\xi)$ linearly between its values at $\xi = \alpha_i$ and $\xi = \alpha_i + h_i$. For $\xi = 0$ we define $\rho(0, \theta) = 0$. The function $\eta(\xi, \theta)$ is defined as follows. We set $\eta(\xi, \theta) = \theta$ for $\xi = \alpha_{i+1} + kh_i$, for each i and each $k = 1, \ldots, 2^i - 1$. For ξ halfway between $\alpha_{i+1} + kh_i$ and $\alpha_{i+1} + (k+1)h_i$ we define $\eta(\xi, \theta) = \theta + \alpha_i$. For the rest of the pairs (ξ, θ) with ξ in the interval $[\alpha_{i+1} + kh_i, \alpha_{i+1} + (k+1)h_i]$ we set $\eta(\xi, \theta)$ to be piecewise linear between the values at the two sides and at the middle of the interval; this for any interval defined by each i and each $k = 1, \ldots, 2^i - 1$. For $\xi = 0$ we define $\eta(0, \theta) = \theta$.

The mapping T given in the above definition is continuous. The choice of h_i, specifically the choice $h_i \alpha_i = 2^i$, with the fact that for fixed ξ the mapping is locally expanding in the θ-coordinate, implies that the images by $T^n(\cdot)$ as $n \to \infty$ of any interval in X of the form $\{(\xi, \theta) : \xi_1 < \xi < \xi_2\}$, are stretched evenly along the circle $\{0\} \times S^1$. Hence the iterates $T^n(\cdot)$ converge as Young measures to the constant-valued Young measure which is the Lebesgue measure on $\{0\} \times S^1$. Yet T is not mixing with respect to the limit measure. Indeed, the subset $\{0\} \times S^1$ consists of fixed points of T. (This does not contradict Theorem 4.4 since the Lebesgue measure on the circle $\{0\} \times S^1$ is orthogonal to the Lebesgue measure on X.) This example also serves the purpose of Example 3.8; but Example 3.8 has the additional property that the empirical measures $\sigma_T^n(x)$ converge individually to the Lebesgue measure on $\{0\} \times S^1$. Such a property cannot hold for iterates.

5. *Examples and comments*

In this section we provide examples and comments illustrating and applying the abstract results of the previous sections.

Example 5.1. Let $X = [0, 1]$ be endowed with the Lebesgue measure dx and let T be the so-called tent mapping; namely, $T(x) = \min(2x, 2(1 - x))$ (see, e.g., Blank [**8**]). A direct inspection reveals that $T^n(x) = T(nx \pmod 1)$. As Young measures in $\mathcal{Y}(([0, 1], dx), [0, 1])$, the iterates $T^n(\cdot)$ converge to the Young measure with a constant value, namely, the Lebesgue measure. By Theorem 4.2 the pair (T, dx) is mixing. By Corollary 4.5 there is no other mixing pair (T, μ) with μ not orthogonal to the Lebesgue measure.

Example 5.2. Consider a variant of the preceding example where $T(x) = 2x \bmod 1$, namely the dyadic map (see, e.g., Blank [**8**]). Again, a direct inspection reveals that

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

$T^n(x) = T(nx \pmod 1)$, and hence the iterates $T^n(\cdot)$ converge as Young measures in $\mathcal{Y}(([0,1], dx), [0,1])$ to the Lebesgue measure. The map T is discontinuous. It is, however, dx-continuous. Hence, by Theorem 4.2 and Corollary 4.5, (T, dx) is mixing and there is no other mixing pair (T, μ) with T being μ-continuous and μ not orthogonal to the Lebesgue measure.

Example 5.3. Consider as a base measure for $[0, 1]$ the Dirac measure δ_0, namely the measure supported on $\{0\}$. For $T(x) = \min(2x, 2(1-x))$ as in Example 5.1, the iterates $T^n(x)$ converge almost everywhere to the constant measure δ_0. The mixing property (which is trivial of course) follows. Example 5.1 combined with this example shows that the non-orthogonality condition in Corollary 4.5 cannot be removed.

Example 5.4. Consider as a base measure for $[0, 1]$ the combination $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{2/3}$, and let T be as in Example 5.1. Then the points in the support of the measure are fixed points of the map T. In particular, the iterates $T^n(\cdot)$ converge pointwise, but not to a constant-valued Young measure. This shows that the property that was used practically throughout the paper, namely that the iterates converge to a constant-valued Young measure, depends heavily on the choice of the underlying measure.

PROPOSITION 5.5. *Let (X, λ) be a probability measure space and let $T : X \to X$ be measurable. Suppose that $\sigma_T^n(\cdot)$ converge as Young measures in $\mathcal{Y}((X, \lambda), X)$ to a Young measure, say $\boldsymbol{\nu}$, with a constant value, say $\nu(x) = \nu$ for λ-almost every x. Then any probability measure μ which is both invariant under T and absolutely continuous with respect to λ coincides with ν.*

Proof. If μ is absolutely continuous with respect to λ, then by Lemma 2.2, the convergence of $\sigma_T^n(\cdot)$ is also valid in $\mathcal{Y}((X, \mu), X)$. If in addition μ is invariant, then the marginal on the image space of each iterate $T^n(\cdot)$, and in particular the marginal of each empirical measure $\sigma_T^n(\cdot)$, is equal to μ. The limit of these marginals is equal to the common value ν, hence $\mu = \nu$. □

Remark 5.6. It may be possible that in the preceding proposition no invariant measure of T exists; this is the case in Example 3.7. It may also be possible that there exist invariant measures that are not absolutely continuous with respect to λ.

Example 5.7. Consider the logistic map on the unit interval $T(x) = 4x(1-x)$. A direct computation reveals that an invariant measure of T is determined by the density $d\mu = \pi^{-1}(x(1-x))^{-1/2}\,dx$, with dx representing the Lebesgue measure. It is also easy to see that the iterates $T^n(\cdot)$ converge as Young measures to a Young measure with a constant value almost everywhere. Since μ and the Lebesgue measure are absolutely continuous with respect to each other, it follows from Corollary 4.5 that the common value of the Young measure limit of iterates is equal to μ. Since T is continuous, it follows from Theorem 4.4 that the pair (T, μ) is mixing, and by Corollary 4.5 it is the unique mixing pair with a measure not orthogonal to the Lebesgue measure. By Proposition 5.5 it is also the unique invariant measure of T that is absolutely continuous with respect to the Lebesgue measure. It is easy to find mixing pairs with a measure orthogonal to the Lebesgue measure,

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

or an invariant measure not orthogonal to, but also not absolutely continuous with respect to the Lebesgue measure.

Example 5.8. Consider the Gauss transformation $T(x) = 1/x - [1/x]$ on the unit interval, where brackets stand for the integer part of the real number (see, e.g., Mañé [**11**]). An invariant measure μ can be computed directly, given by the density $d\mu = (1/\log 2)(1+x)^{-1}\,dx$, with dx again representing the Lebesgue measure. It is also not difficult to see that the iterates $T^n(\cdot)$ converge as Young measures to a Young measure of a constant value. Employing the μ-continuity of T, we deduce (similar to the previous example) from Theorem 4.4 and Corollary 4.5 that (T, μ) is mixing, and that it is the unique mixing pair not orthogonal to the Lebesgue measure. From Proposition 5.5 we deduce that μ is the unique invariant measure that is absolutely continuous with respect to the Lebesgue measure. This consequence is derived in Mañé [**11**, Ch. III, Theorem 1.2] from Bowen's results on Markov transformations.

Acknowledgements. ZA is Incumbent of the Hettie H. Heineman Professorial Chair in Mathematics. His research is supported by grants from the Israel Science Foundation, the MINERVA Foundation, Germany, and by INTAS, Belgium. MG's research was supported by a grant from the Leverhulme Foundation, and by the Belfer Institute for Mathematics and Computer Science at the Weizmann Institute.

References

[1] Z. Artstein. Invariant measures of differential inclusions applied to singular perturbations. *J. Diff. Eqns* **152** (1999), 289–307.

[2] Z. Artstein. Invariant measures of set-valued maps. *J. Math. Anal. Appl.* **252** (2000), 696–709.

[3] Z. Artstein and A. Vigodner. Singularly perturbed ordinary differential equations with dynamic limits. *Proc. R. Soc. Edinburgh* A **126** (1996), 541–569.

[4] Z. Artstein and R. J-B Wets. Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM J. Optim.* **4** (1994), 537–550.

[5] J. M. Ball. A version of the fundamental theorem for Young measures. *Partial Differential Equations and Continuum Models of Phase Transitions (Lecture Notes in Physics, 344)*. Eds. M. Ruscle, D. Serre and M. Slemrod. Springer, Berlin, 1989, pp. 207–215.

[6] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete Time Case*. Academic, New York, 1978.

[7] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[8] M. Blank. *Discreteness and Continuity in Problems of Chaotic Dynamics (Translations of Monographs, 161)*. American Mathematical Society, Providence, RI, 1997.

[9] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms (Lecture Notes in Mathematics, 470)*. Springer, Berlin, 1975.

[10] M. Grinfeld and P. A. Knight. Young measures and invariance. *Proc. EQUADIFF '99*. Vol. 2, World Scientific, Singapore, 2000, pp. 1039–1041.

[11] R. Mañé. *Ergodic Theory and Differentiable Dynamics*. Springer, Berlin, 1987.

[12] W. Miller and E. Akin. Invariant measures for set-valued dynamical systems. *Trans. Amer. Math. Soc.* **351** (1999), 1203–1225.

[13] P. Pedregal. Optimization, relaxation and Young measures. *Bull. Amer. Math. Soc. (N.S.)* **36** (1999), 27–58.

[14] L. Tartar. Compensated compactness and applications in partial differential equations. *Nonlinear Analysis and Mechanics*. Ed. R. J. Knops. Pitman, London, 1979, pp. 136–212.

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press

[15] F. Topsøe. Preservation of weak convergence under mappings. *Ann. Math. Statist.* **38** (1967), 1661–1665.
[16] A. Vershik. Many-valued measure-preserving mappings (polymorphisms) and Markovian operators. *J. Soviet Math.* **23** (1983) 2243–2266.
[17] J. Warga. *Optimal Control of Differential and Functional Equations*. Academic, New York, 1976.
[18] L. C. Young. *Lectures on the Calculus of Variations and Optimal Control Theory*. Saunders, New York, 1969.

https://doi.org/10.1017/S0143385702000731 Published online by Cambridge University Press