Ergod. Th. & Dynam. Sys., page 1 of 16
doi:10.1017/etds.2023.119

Lifting generic points

TOMASZ DOWNAROWICZ† and BENJAMIN WEISS‡

† *Faculty of Pure and Applied Mathematics, Wrocław University of Technology, Wrocław, Poland*
(e-mail: downar@pwr.edu.pl)

‡ *Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel*
(e-mail: weiss@math.huji.ac.il)

(*Received* 6 *September* 2023 *and accepted in revised form* 4 *December* 2023)

Abstract. Let (X, T) and (Y, S) be two topological dynamical systems, where (X, T) has the weak specification property. Let ξ be an invariant measure on the product system $(X \times Y, T \times S)$ with marginals μ on X and ν on Y, with μ ergodic. Let $y \in Y$ be quasi-generic for ν. Then there exists a point $x \in X$ generic for μ such that the pair (x, y) is quasi-generic for ξ. This is a generalization of a similar theorem by T. Kamae, in which (X, T) and (Y, S) are full shifts on finite alphabets.

1. *Introduction*

All terminology used freely in this introduction is explained in the preliminaries (§2).

Let $\pi : (X, T) \to (Y, S)$ be an extension of a compact dynamical system (Y, S) and suppose that ν is an ergodic measure for S. This measure can always be lifted to an invariant measure on X (by the Hahn–Banach theorem). It then follows that there exists an ergodic measure μ that projects to ν. Clearly, any generic point for μ will project to a generic point for ν. It is natural to ask whether all ν-generic points can be obtained in this way. In other words: does every ν-generic point have a μ-generic lift? It is not difficult to show that if μ is a unique lift of ν then the answer is yes. In fact, in this case if $y \in Y$ is generic for ν then any $x \in \pi^{-1}(y)$ will be generic for μ. However, if the extension of ν is not unique then the answer might be negative. Such examples can be obtained as follows.

Consider a minimal almost one-to-one extension $\pi : X \to Y$ where Y is strictly ergodic but X is not (cf. Furstenberg and Weiss [**FW**] for examples of such systems). Then all invariant measures on X project to the unique measure ν on Y. In this situation all points in Y are generic for ν. Now, let $y \in Y$ be a point with a unique preimage $x \in X$ (by assumption such a point exists). Then either x is not generic for any invariant measure on X or it is generic for one such measure (say, μ_0). In either case, there exists an invariant measure μ_1 on X such that x is not generic for μ_1. Thus, the point y generic for ν does not lift to a point generic for the lift μ_1 of ν.

However, this example does not provide an answer to a more subtle question: does every ν-generic point have a generic lift (without specifying for which measure extending ν)? In general, the answer to such a relaxed question is also negative. We will show this using the same example as before and the following theorem.

THEOREM 1.1. *Let (X, T) be a topological dynamical system with (at least) two different ergodic measures μ and ν, both having full topological support. Then there exist a dense G_δ-set $B \subset X$ and a continuous function f on X such that for any $x \in B$ the ergodic averages*

$$A_n(f, x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)$$

oscillate.

Proof. Since the measures differ, there exists a continuous function f on X whose integral with respect to μ is greater than one while its integral with respect to ν is less than zero. Now, for a natural number N we define

$$E_N = \left\{ x \in X : \text{ there exists}_{\,n>N} \ \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) > 1 \right\}.$$

This set is clearly open, and it is dense since the generic points for μ are dense. Define a similar set F_N replacing '> 1' by '< 0'. Then the desired set B is the countable intersection

$$B = \bigcap_{N \geq 1} (E_N \cap F_N).$$

By the Baire theorem, this set is a dense G_δ, and clearly no $x \in B$ is generic for any measure. □

Now let us go back to the example. Since the system (X, T) is minimal, all its invariant measures have full topological support. By Theorem 1.1, there is a dense G_δ-set B of points which are not generic for any measure. As in any minimal almost one-to-one extension, the 'singleton fibers' (that is, points which are unique preimages of their images) also form a dense G_δ-set (call it A) in X. Then the intersection $A \cap B$ is non-empty and any point in its image is generic (for ν) but has no generic lift.

An even more extreme situation occurs in yet another example. In [**F**], Furstenberg constructs a non-uniquely ergodic minimal skew product on the 2-torus over an irrational rotation, where the fiber maps are also rotations. Since rotation by a fixed angle on the fibers commutes with the skew product, in every fiber either each point is generic for some invariant measure or none of them is. Theorem 1.1 implies that there are points not generic for any measure (actually, this property is explicit in [**F**]—this is how Furstenberg showed non-unique ergodicity of the skew product). It follows that there are entire fibers (circles) with no generic points even though in the base all points are generic.

Before we discuss a positive result we need to mention two important issues. The first one is the phenomenon of *quasi-generating* invariant measures, that is, generating them along a subsequence of averages. Replacing the term 'generic' by 'quasi-generic' may lead to either stronger or weaker results, depending on where the replacement is done (in the assumption or in the thesis). The second issue is a specific way of extending a system by *joining it* with another system. Any extension can be viewed practically as a joining of the system with its extension (the joining is then supported by the graph of the factor map), but it is often essential to know that the extension can be obtained as a joining with a system having some specific properties (ergodicity, specification property, etc.) which the entire extension does not necessarily enjoy.

In the early 1970s, Teturo Kamae studied normal sequences and the phenomenon of normality-preserving subsequences. In symbolic dynamics a sequence over a finite alphabet is *normal* if it is generic for the uniform Bernoulli measure. An increasing subsequence of natural numbers $y = (n_k)_{k\geq 1}$ *preserves normality* if $x|_y = (x_{n_k})_{k\geq 1}$ is normal for any normal sequence $(x_n)_{n\geq 1}$. A few years earlier, Weiss [**W**] proved that subsequences of positive lower density which are completely deterministic preserve normality. A subsequence y is *completely deterministic* if its indicator function, viewed as an element of the shift on two symbols, quasi-generates only measures of entropy zero. Kamae [**K**] proved the opposite implication: *only* completely deterministic subsequences preserve normality. Given a non-deterministic subsequence y (that is, one that quasi-generates some measure ν of positive entropy), he needed to find a normal sequence x such that $x|_y$ is not normal. Skipping the details, let us just say that he needed to 'pair' the subsequence y with a normal (that is, generic for the uniform Bernoulli measure λ) sequence x, such that the pair (x, y) is generic for a specific joining ξ of λ and ν. In order to do so, he proved a more general theorem, which motivates our current work. We take the liberty of rephrasing the statement in the language that we use throughout this paper.

THEOREM 1.2. [**K**] *Let ξ be a joining of two invariant measures, μ and ν, supported on symbolic systems $\Lambda_1^{\mathbb{N}}$ and $\Lambda_2^{\mathbb{N}}$, respectively (Λ_1 and Λ_2 are finite alphabets). Let $y \in \Lambda_2^{\mathbb{N}}$ be quasi-generic for ν, that is, it generates ν along a subsequence of averages indexed by $\mathcal{J} = (n_k)_{k\geq 1}$. Then there exists $x \in \Lambda_1^{\mathbb{N}}$ such that the pair (x, y) generates ξ along $\mathcal{J}$. If μ is ergodic then x can be chosen generic for μ.*

This theorem found another application in the work of Rauzy [**R**], who studied normality preservation in a different sense. Let us identify all real numbers with their expansions in some fixed base $b \geq 2$. A real number is called *normal (in base b)* if its expansion is a normal sequence. A real number y *preserves normality* if $x + y$ is normal for any normal number x. Rauzy proved that a number y preserves normality if and only if the expansion of y is completely deterministic.

Notice that Theorem 1.2 is actually very strong. First of all, it applies to any situation when a 'symbolic' measure ν is lifted to a 'symbolic' measure ξ. Also note that ν is not assumed ergodic, it suffices that it admits a quasi-generic point y (which is always true within a full shift). If $\mathcal{J} = \mathbb{N}$ then y is simply generic for ν and the theorem allows it to be lifted to a pair (x, y) generic for ξ. Even when $\mathcal{J}$ is an essential subsequence (and there is no hope of making the lift (x, y) generic), as soon as μ is ergodic, the point x 'paired' with y still can be generic rather than just quasi-generic. The only weakness of the theorem is that x is found within the *full shift* $\Lambda_1^{\mathbb{N}}$, even when μ is supported by a proper subshift. In other words, the theorem does not allow y to be lifted within an *a priori* given topological (symbolic) extension.

Our paper focuses exactly on this problem. Our goal is to find conditions under which the 'paired' point x (generic for μ) can be found within the *a priori* given topological system (X, T) being joined with (Y, S). The conditions turn out to be ergodicity of μ (like in the original theorem), and the *weak specification property* of (X, T). We prove the following theorem.

THEOREM 1.3. *Let (X, T) and (Y, S) be topological dynamical systems and let ξ be an invariant measure on the product system $(X \times Y, T \times S)$ with marginals μ and ν on X and Y, respectively. Assume that the system (X, T) has the weak specification property and that μ is ergodic under T. Suppose also that $y \in Y$ is quasi-generic for the measure ν. Then there exists a point $x \in X$, generic for μ, such that the pair (x, y) is quasi-generic for ξ.*

Let us mention that the weak specification property is satisfied by many systems such as ergodic mixing Markov shifts, ergodic toral automorphisms, and in fact any endomorphisms of compact Abelian groups for which the Haar measure is ergodic (see [**D**] and use natural extension). An advantage of our result is that it is not restricted to symbolic systems and that x is found within the space X. A disadvantage is that the pair (x, y) is only quasi-generic for ξ, even when $\mathcal{J} = \mathbb{N}$.

The strength of Theorem 1.3 lies in the possibility of lifting *any* generic point (not just almost any) y to a pair (x, y) quasi-generic for ξ. If ξ is ergodic, such a possibility for ν-almost all y is a trivial fact. Thus, the theorem can be useful when topological, rather than measure-theoretic, precision is required.

A concrete application of Theorem 1.3 occurs in the forthcoming paper [**BD**], where Rauzy's equivalence between normality preservation and determinism is generalized to a wider context. That is to say, the following problem is addressed.

Question 1.4. Let $T : X \to X$ be a surjective endomorphism of a compact metrizable Abelian group, such that the Haar measure λ on X is T-ergodic and has finite entropy. Let us call a point $x \in X$ *normal* if it is generic for λ. Is it true that y *preserves normality* (that is, $x + y$ is normal for any normal $x \in X$) if and only if y is completely deterministic?

In [**BD**] we prove sufficiency relatively easily, but the harder direction (necessity) is shown only for selected groups X (tori, solenoids, and countable direct products $\bigoplus_{n\geq 1} \mathbb{Z}_d$ $(d \geq 2)$; the necessity in full generality remains open). In all these cases Theorem 1.3 plays a crucial role in the proofs.

Our paper is organized as follows. Section 2 contains all necessary definitions and notational conventions. In §3 we provide three key lemmas together with auxiliary propositions needed in their proofs. The propositions are quite standard while the lemmas may be considered of independent interest. Finally, in §4 we present the proof of Theorem 1.3.

2. *Preliminaries*

Let (X, T) be a topological dynamical system, where X is a compact metric space and T is a continuous surjection. By $\mathcal{M}(X)$ we will denote the set of all Borel probability measures on X. Since no other measures will be considered, the elements of $\mathcal{M}(X)$ will be called *measures* for short. By $\mathcal{M}_T(X)$ we will denote the subset of $\mathcal{M}(X)$ containing all measures that are T-invariant, that is, such that $\mu(T^{-1}A) = \mu(A)$ for all Borel sets $A \subset X$. When the transformation T is fixed, the elements of $\mathcal{M}_T(X)$ will be called *invariant measures*. The sets $\mathcal{M}(X)$ and $\mathcal{M}_T(X)$ are equipped with the weak* topology, which makes both these sets compact convex and metrizable with a convex metric. (By definition, a sequence $(\mu_n)_{n\geq 1}$ of measures converges in the weak* topology to a measure μ if, for any continuous (real or complex) function f on X, the integrals $\int f\,d\mu_n$ converge to $\int f\,d\mu$. One of the standard convex metrics compatible with this topology is given by

$$d(\mu, \nu) = \sum_{n=1}^{\infty} 2^{-n}\bigg|\int f_n\,d\mu - \int f_n\,d\nu\bigg|,$$

where $(f_n)_{n\geq 1}$ is a sequence of continuous functions on X with values in the interval [0, 1], linearly dense in the space $C(X)$ of all continuous real functions on X.) It is well known that the extreme points of $\mathcal{M}_T(X)$ are precisely the *ergodic measures*, that is, invariant measures μ such that $\mu(A \triangle T^{-1}A) = 0 \implies \mu(A) \in \{0, 1\}$, for any Borel set $A \subset X$.

We will be using the following notation. For two integers $a \leq b$, we will denote by $[a, b]$ the interval of integers $\{a, a+1, a+2, \ldots, b\}$. Given a point $x \in X$ and $0 \leq a \leq b$, we denote by $x[a, b]$ the ordered finite segment of the orbit of x,

$$x[a, b] = (T^a x, T^{a+1}x, \ldots, T^b x),$$

while by $\mu_{x[a,b]}$ we will understand the normalized counting measure on $x[a, b]$,

$$\mu_{x[a,b]} = \frac{1}{b-a+1}\sum_{n=a}^{b} \delta_{T^n x}$$

(here δ_x denotes the Dirac measure concentrated at x). We will call this measure the *empirical measure* associated with the orbit segment.

A point x is said to *quasi-generate* (or *be quasi-generic for*) a measure μ if μ is an accumulation point of the sequence of measures $(\mu_{x[0,n]})_{n\geq 1}$ (any such measure μ is invariant). In this case there exists an increasing sequence of natural numbers $\mathcal{J} = (n_k)_{k\geq 1}$ such that $\lim_{k\to\infty} \mu_{x[0,n_k]} = \mu$. We will say that x *generates* μ *along* $\mathcal{J}$. If the sequence $(\mu_{x[0,n]})_{n\geq 1}$ converges to μ then we say that x *generates* (or *is generic for*) μ (in other words, 'generic' equates to 'generic along $\mathbb{N}$'). It follows from the pointwise ergodic theorem that every ergodic measure μ possesses generic points (in fact μ-almost all points are such). The following obvious fact will be used several times.

Remark 2.1. Two increasing sequences of natural numbers (say, $(n_k)_{k\geq 1}$ and $(m_k)_{k\geq 1}$) will be called *equivalent* if $\lim_{k\to\infty}(n_k/m_k) = 1$. It is obvious that the upper (and lower) densities of any subset of $\mathbb{N}$ evaluated along equivalent sequences are the same. If a point x generates a measure μ along a sequence $(n_k)_{k\geq 1}$ then it generates μ along any sequence $(m_k)_{k\geq 1}$ equivalent to $(n_k)_{k\geq 1}$.

Other key notions in this paper are those of a specification and the specification property.

Definition 2.2

(1) Consider a (finite or infinite) sequence of non-negative integers:

$$a_1 \le b_1 < a_2 \le b_2 < \cdots < a_{N_1} \le b_{N_1} \quad \text{where } N_1 \in \mathbb{N}, \text{ or}$$
$$a_1 \le b_1 < a_2 \le b_2 < a_3 \le b_3 < \cdots .$$

Let $D = \bigcup_N [a_N, b_N]$ (where N ranges over either $[1, N_1]$ or $\mathbb{N}$). By a *specification* with domain D we will mean any function

$$\mathcal{S} : D \to X$$

such that for each N there exists a point x_N such that for each $n \in [a_N, b_N]$ we have

$$\mathcal{S}(n) = T^n(x_N).$$

Since T is surjective, we can equivalently demand that $\mathcal{S}(n) = T^{n-a_N}(x_N)$.

(2) By $\mathcal{S}[a_N, b_N]$ we mean the ordered tuple $(\mathcal{S}(a_N), \mathcal{S}(a_N + 1), \ldots, \mathcal{S}(b_N))$ which equals $x_N[a_N, b_N]$ (or $x_N[0, b_N - a_N]$).

(3) The numbers $l_N = b_N - a_N + 1$ and $g_N = a_{N+1} - b_N - 1$ will be called the *orbit segment lengths* and *gaps* of the specification, respectively.

(4) If D is finite then the *empirical measure* associated with $\mathcal{S}$ is defined as

$$\mu_{\mathcal{S}} = \frac{1}{|D|} \sum_{n \in D} \delta_{\mathcal{S}(n)}.$$

(5) An infinite specification $\mathcal{S}$ *generates* (or *is generic for*) a measure μ along a sequence $\mathcal{J} = (n_k)_{k \ge 1}$ if the measures associated to the specification $\mathcal{S}$ restricted to $D \cap [0, n_k]$ converge to μ (in general, μ need not be invariant).

(6) We say that (the orbit of) a point $x \in X$ ε-shadows the specification $\mathcal{S}$ if

$$\text{for all } n \in D, \quad d(\mathcal{S}(n), T^n(x)) < \varepsilon.$$

(7) A system (X, T) has the *weak specification property* if for every $\varepsilon > 0$ there exists a function $M_\varepsilon : \mathbb{N} \to \mathbb{N}$ satisfying $\lim_{l\to\infty}(M_\varepsilon(l)/l) = 0$, such that any finite specification (with any finite number N_1 of orbit segments) satisfying, for each $N \in [1, N_1]$, the inequality $g_N \ge M_\varepsilon(l_{N+1})$ is ε-shadowed by an orbit.

The last condition asserts, roughly speaking, that any appropriately spaced finite sequence of orbit segments (where each gap is adjusted to the length of the following segment, according to the function M_ε) can be ε-shadowed by a single orbit.

3. *Preparatory statements*

The proof of Theorem 1.3 relies on three key lemmas. The first one is concerned with increasingly good shadowing of certain infinite specifications.

LEMMA 3.1. *Let (X, T) be a topological dynamical system with the weak specification property with a family of functions $\{M_\varepsilon : \varepsilon > 0\}$. Let $(\varepsilon_k)_{k\ge 0}$ be a summable sequence of positive numbers. Let $D = \bigcup_{N=1}^{\infty} [a_N, b_N]$ and let $\mathcal{S} : D \to X$ be an infinite specification* *satisfying, for some increasing sequence* $(N_k)_{k\ge 0}$ *of non-negative integers starting with* $N_0 = 0$*, the following condition: for each* $k \ge 1$ *and all* $N \in [N_{k-1}+1, N_k]$ *we have*

$$g_N \ge M_{\varepsilon_k}(l_{N+1}). \tag{3.1}$$

Then there exists a point x_0 *such that*

$$\lim_{n\in D} d(\mathcal{S}(n), T^n x_0) = 0. \tag{3.2}$$

Proof. Let $\mathcal{S}_1$ denote the specification $\mathcal{S}$ restricted to the initial N_1 orbit segments. This finite specification satisfies the inequality $g_N \ge M_{\varepsilon_1}(l_{N+1})$, hence it can be ε_1-shadowed by the orbit of some point $x_1 \in X$.

We continue by induction. Suppose that we have found a point $x_k \in X$ which satisfies

$$\text{for all } i \in [1,k],\ \text{for all } N \in [N_{i-1}+1, N_i],\ \text{for all } n \in [a_N, b_N], \quad d(\mathcal{S}(n), T^n x_k) \le \sum_{j=i}^{k} \varepsilon_j. \tag{3.3}$$

We define a new specification $\mathcal{S}_{k+1}$ on

$$[0, b_{N_k}] \cup \bigcup_{N=N_k+1}^{N_{k+1}} [a_N, b_N]$$

as follows. We let $\mathcal{S}_{k+1}[0, b_{N_k}] = x_k[0, b_{N_k}]$, while for $N \in [N_k+1, N_{k+1}]$ we let $\mathcal{S}_{k+1}[a_N, b_N] = \mathcal{S}[a_N, b_N]$. It will be convenient not to change the enumeration of the orbit segments (except for the first one, which is new), and of the gaps (the first gap of $\mathcal{S}_{k+1}$ coincides with the N_kth gap of $\mathcal{S}$). Then $\mathcal{S}_{k+1}$ satisfies $g_N \ge M_{\varepsilon_k}(l_{N+1})$ for all $N \in [N_k, N_{k+1}-1]$ (that is, for all gaps of $\mathcal{S}_{k+1}$), hence it can be ε_k-shadowed by the orbit of some point $x_{k+1} \in X$. It is clear that x_{k+1} satisfies (3.3) with the parameter $k+1$ in place of k. This concludes the induction. We let x_0 be any accumulation point of the sequence $(x_k)_{k\ge 1}$. As easily seen, this point satisfies, for all $n \in D$, the inequality

$$d(\mathcal{S}(n), T^n(x_0)) \le \sum_{j=k_n+1}^{\infty} \varepsilon_j,$$

where k_n is the unique integer $k \ge 0$ such that $n \in [a_N, b_N]$ with $N \in [N_k+1, N_{k+1}]$. Since the sums on the right-hand side are tails of a convergent series, these distances tend to zero, as claimed. □

Remark 3.2. It is easily seen that if the domain D of $\mathcal{S}$ in the above lemma has density one then the point x_0 from that lemma quasi-generates the same invariant measures as $\mathcal{S}$.

The second key lemma requires two rather standard propositions from convex analysis. Although they are well known to specialists, it is hard to find them in the exact formulation. Thus, we provide them with proofs.

Let $(\mathcal{M}, d)$ be a compact convex set in a locally convex metric space (the reader may think of $(\mathcal{M}(X), d)$, where d is some standard metric compatible with the weak* topology). The elements of $\mathcal{M}$ will be denoted by the letters μ, ν.

PROPOSITION 3.3. *Let $T : \mathcal{M} \to \mathcal{M}$ be a continuous affine transformation. Then the set $\mathcal{M}_T \subset \mathcal{M}$, consisting of T-invariant elements, is non-empty and for any $\varepsilon > 0$ there exists $n_\varepsilon \geq 1$ such that, for any $\nu \in \mathcal{M}$ and any $n \geq n_\varepsilon$, we have $d((1/n)\sum_{i=0}^{n-1} T^i(\nu), \mathcal{M}_T) < \varepsilon$.*

Proof. We can assume that $\mathsf{diam}(\mathcal{M}) = 1$. Denote $A_n(\nu) = (1/n)\sum_{i=1}^{n-1} T^i(\nu)$. Then, by convexity of the metric and diameter 1 of $\mathcal{M}$, we easily see that

$$d(T(A_n(\nu)), A_n(\nu)) \leq \frac{1}{n}.$$

This in turn implies that any limit point of any sequence of the form $A_n(\nu_n)$ (with $\nu_n \in \mathcal{M}$) is T-invariant. Such limit points exist by compactness, hence we get that $\mathcal{M}_T \neq \emptyset$. Suppose that the second part of the proposition does not hold. This means that there exist $\varepsilon > 0$ and an increasing sequence $(n_k)_{k\geq 1}$ of natural numbers, and a sequence $(\nu_k)_{k\geq 1}$ of points of $\mathcal{M}$, such that $d(A_{n_k}(\nu_k), \mathcal{M}_T) \geq \varepsilon$ for all $k \geq 1$. But we have just proved that all accumulation points of the sequence $(A_{n_k}(\nu_k))_{k\geq 1}$ belong to $\mathcal{M}_T$, so we have a contradiction. □

Recall that if ξ is a probability measure on $\mathcal{M}$ then there exists a unique point $\mu \in \mathcal{M}$, called the *barycenter* of ξ, such that for every affine continuous function f one has

$$f(\mu) = \int f(\nu)\, d\xi(\nu).$$

The barycenter map is denoted by either $\xi \mapsto \mathsf{bar}(\xi)$ or by $\xi \mapsto \int \nu\, d\xi(\nu)$ (the integral in the sense of Pettis). It is well known that if the set of all probability measures on $\mathcal{M}$ is endowed with the weak* topology then the barycenter map $\xi \mapsto \mathsf{bar}(\xi)$ is continuous. In the next proposition, the reader may think of $\mathcal{M}$ representing $\mathcal{M}_T(X)$ in a dynamical system (X, T), and μ representing an ergodic measure.

PROPOSITION 3.4. *Let μ be an extreme point of $\mathcal{M}$. Then for any $\varepsilon > 0$ there exists $\delta > 0$ such that, whenever a probability measure ξ on $\mathcal{M}$ satisfies $d(\mathsf{bar}(\xi), \mu) < \delta$, we have*

$$\xi\{\nu \in \mathcal{M} : d(\mu, \nu) < \varepsilon\} > 1 - \varepsilon.$$

Proof. If the statement is false then there exists a sequence of measures $(\xi_k)_{k\geq 1}$ on $\mathcal{M}$ such that $\lim_{k\to\infty} \mathsf{bar}(\xi_k) = \mu$ and $\xi_k\{\nu \in \mathcal{M} : d(\mu, \nu) < \varepsilon\} \leq 1 - \varepsilon$ for each $k \geq 1$. Since the function which associates to a measure ξ the value $\xi(U)$, where U is an open set, is lower semicontinuous in the weak* topology, we get that if ξ is an accumulation point of the sequence $(\xi_k)_{k\geq 1}$ then $\xi\{\nu \in \mathcal{M} : d(\mu, \nu) < \varepsilon\} \leq 1 - \varepsilon$. On the other hand, by continuity of the barycenter map, we have $\mathsf{bar}(\xi) = \mu$. Since μ is an extreme point of $\mathcal{M}$, the only measure on $\mathcal{M}$ with barycenter at μ is the Dirac measure δ_μ. We conclude that $\xi = \delta_\mu$. This is a contradiction, since $\delta_\mu\{\nu \in \mathcal{M} : d(\mu, \nu) < \varepsilon\} = 1$. □

We proceed with the second key lemma needed in the proof of Theorem 1.3.

LEMMA 3.5. *Let μ be an ergodic measure on a topological dynamical system (X, T) which has the weak specification property. Let $x_0 \in X$ be quasi-generic for μ and let $\mathcal{J} = (n_k)_{k\geq 1}$ be a sequence along which x_0 generates μ. Then there exist a point $\bar{x}_0 \in X$*

generic for μ and a set $\mathbb{M} \subset \mathbb{N}$ of upper density one achieved along a subsequence of $\mathcal{J}$, such that

$$\lim_{n\in\mathbb{M}} d(T^n\bar{x}_0, T^n x_0) = 0.$$

Proof. We start by fixing a summable sequence of positive numbers $(\varepsilon_k)_{k\geq 1}$. In view of Lemma 3.1 and Remark 3.2, it suffices to construct a specification $\mathcal{S}$ on a domain D satisfying the following four conditions:

(1) the assumptions of Lemma 3.1;
(2) the domain D of $\mathcal{S}$ has density one;
(3) $\lim_{n\in\mathbb{M}} d(\mathcal{S}(n), T^n(x_0)) = 0$, where $\mathbb{M} \subset D$ has upper density one achieved along a subsequence of $\mathcal{J}$;
(4) $\mathcal{S}$ is generic for μ.

We choose a sequence of positive integers $(l_k)_{k\geq 0}$. The sequence should grow so fast that the ratios $M_{\varepsilon_k}(l_k)/l_k$ are all smaller than one and tend to zero. For each $k \geq 1$ we let $L_k = l_k + M_{\varepsilon_k(l_k)}$. Next, we replace $\mathcal{J}$ by a fast-growing subsequence and from now on $\mathcal{J} = (n_k)_{k\geq 1}$ will denote that subsequence. Initially we require that the ratios l_k/n_k and n_k/n_{k+1} tend to zero as k grows. More conditions on the speed of growth of the sequences $(l_k)_{k\geq 1}$ and $(n_k)_{k\geq 1}$ will be specified later.

The specification $\mathcal{S}$ is created in three steps. The first auxiliary specification $\mathcal{S}'$ is just a partition of the orbit of x_0 without gaps. We begin by partitioning it into segments of length L_1 until we cover the coordinate n_1. Then we continue by partitioning the remaining part of the orbit of x_0 into segments of length L_2 until we cover the coordinate n_2 and so on. To be precise, we create segments $\mathcal{S}'[a_N, b_N] = x_0[a_N, b_N]$ (where $N \geq 1$) satisfying:

(i) $a_1 = 0$;
(ii) for $N \geq 2$, $a_N = b_{N-1} + 1$;
(iii) $b_N = a_N + L_k - 1$, for $N \in [N_{k-1} + 1, N_k]$, where
(iv) for each $k \geq 1$, N_k is such that $n_k \in [a_{N_k}, b_{N_k}]$

(for consistency of notation, we have let $N_0 = 0$). It is elementary to see that, since the ratios $M_{\varepsilon_k}(l_k)/l_k$ and l_k/n_k tend to zero, the ratios b_{N_k}/n_k tend to one. Thus, by Remark 2.1, the point x_0 generates μ along the sequence $(b_{N_k})_{k\geq 1}$. From now on, we redefine the sequence $\mathcal{J}$ to be $(b_{N_k})_{k\geq 1}$ (and let $n_k = b_{N_k}$; we also let $n_0 = 0$). This new sequence still satisfies $n_k/n_{k+1} \to 0$.

The empirical measures $\mu_{x_0[0,n_k]}$ tend to μ. Since n_{k-1} is eventually negligible in comparison with n_k, the following statement holds:

$$\text{the empirical measures } \mu_{x_0[n_{k-1}+1,n_k]} \text{ tend to } \mu \text{ as } k \text{ grows.} \tag{3.4}$$

The second auxiliary specification $\mathcal{S}''$ is obtained from $\mathcal{S}'$ by truncating all orbit segments (except the first) on the left, to allow for future shadowing. More precisely, we let $a'_1 = a_1 = 0$ and for any $k \geq 1$ and any $N \in [N_{k-1} + 1, N_k]$ (except for $N = 1$) we let $a'_N = a_N + M_{\varepsilon_k}(l_k)$ (since $M_{\varepsilon_k}(l_k) < l_k$, we have $a'_N < b_N$). Then on the new domain

$$D = \bigcup_{N\geq 1} [a'_N, b_N],$$

we define the specification $\mathcal{S}''$ by $S''[a'_N, b_N] = x_0[a'_N, b_N]$. This new specification has, for $N \in [N_{k-1}+1, N_k]$ (except for $N = 1$), orbit segments of length l_k preceded by gaps of size $M_{\varepsilon_k}(l_k)$ (the first orbit segment has length L_1 and no preceding gap).

It should be quite obvious that the lower density of D is achieved along the sequence $b_{N_k} + M_{\varepsilon_{k+1}(l_{k+1})}$ (this is the end of the first gap, larger than all preceding gaps). Because the ratios $M_{\varepsilon_k}(l_k)/l_k$ tend to zero, by choosing the numbers n_k (and hence b_{N_k}) sufficiently large in comparison with $M_{\varepsilon_{k+1}}(l_{k+1})$, we can arrange that the density of D equals one, as required in (2).

We now have to go back to the choice of the sequences (l_k) and (n_k) and impose more conditions on the speed of their growth. We select numbers $\delta_k \le \varepsilon_k$ according to Proposition 3.4 with respect to the numbers ε_k and the ergodic measure μ in the role of the extreme point of the compact convex set $\mathcal{M}_T(X)$. If the numbers l_k are (*a priori*) chosen large enough, using Proposition 3.3, we can arrange that

- the empirical measures $\mu_{x_0[a'_N,b_N]}$ with $N \in [N_{k-1}+1, N_k]$ are $\delta_k/3$-close to some invariant measures henceforth denoted by μ_N.

Also, by imposing fast enough growth of the numbers n_k, we may achieve that:

- the empirical measure $\mu_{x_0[n_{k-1}+1,n_k]}$ is $(\delta_k/3)$-close to μ (see (3.4));
- the empirical measures $\mu_{x_0[a_N,b_N]}$ with $N \in [N_{k-1}+1, N_k]$ are $(\delta_k/3)$-close to the respective empirical measures $\mu_{x_0[a'_N,b_N]}$ (and hence $\frac{2}{3}\delta_k$-close to μ_N).

Clearly, the empirical measure $\mu_{x_0[n_{k-1}+1,n_k]} = \mu_{x_0[a_{N_{k-1}+1},b_{N_k}]}$ equals the arithmetic average of the measures $\mu_{x_0[a_N,b_N]}$ with $N \in [N_{k-1}+1, N_k]$. By convexity of the metric, μ is δ_k-close to the arithmetic average of the invariant measures μ_N with $N \in [N_{k-1}+1, N_k]$. By Proposition 3.4, the vast majority of the invariant measures μ_N are ε_k-close to μ, and hence the corresponding empirical measures $\mu_{x_0[a'_N,b_N]}$ are $2\varepsilon_k$-close to μ (we are using $\delta_k < \varepsilon_k$). More precisely, there are fewer than $\varepsilon_k(N_k - N_{k-1})$ parameters $N \in [N_{k-1}+1, N_k]$ (we will call them 'bad'), for which the measure $\mu_{x_0[a'_N,b_N]}$ is not $2\varepsilon_k$-close to μ.

We can now perform the third step in creating the specification $\mathcal{S}$. This is done by replacing in $\mathcal{S}''$ the segments $x_0[a'_N, b_N]$ corresponding to 'bad' parameters $N \in [N_{k-1}, N_k]$ by orbit segments (of the same length) whose associated measures are $2\varepsilon_k$-close to μ. For example, we can choose one 'good' parameter N (there exists such an N) and use the corresponding segment $x_0[a'_N, b_N]$ everywhere we need to make a replacement. This concludes the construction of $\mathcal{S}$.

We need to verify that $\mathcal{S}$ satisfies the desired four properties.

(1) It is clear that the specification $\mathcal{S}$ was created in accordance with the assumptions of Lemma 3.1.

(2) As we have already remarked, the domain D has density one.

(3) Note that $\mathcal{S}''$ agrees with the orbit of x_0 on D (which has density one). Then $\mathcal{S}$ differs from $\mathcal{S}''$ on a set whose frequency in the interval $[n_{k-1}+1, n_k]$ is at most ε_k. Thus the set of the integers n for which $\mathcal{S}(n) \ne \mathcal{S}''(n)$ (or $\mathcal{S}(n)$ is not defined) has lower density zero achieved along the sequence $\mathcal{J}$. The complementary set $\mathbb{M}$ has upper density one achieved along $\mathcal{J}$, and on this set we have $\mathcal{S}(n) = T^n(x_0)$ (which trivially implies the required condition $\lim_{n\in\mathbb{M}} d(\mathcal{S}(n), T^n x_0) = 0$).

(4) Consider a long initial segment of $\mathcal{S}$ (say, $\mathcal{S}|_{[1,n]\cap D}$), and let k be such that $N \in [N_{k-1}+1, N_k]$, where N is determined by the inclusion $n \in [a_N, b_N]$. Then $\mathcal{S}|_{[1,n]\cap D}$ consists essentially of segments of two lengths: l_{k-1}, whose associated empirical measures are $2\varepsilon_{k-1}$-close to μ; and l_k, whose associated empirical measures are $2\varepsilon_k$-close to μ (in either case we have $2\varepsilon_{k-1}$-closeness). This closeness need not apply to the initial part left of the coordinate n_{k-1}, and to the terminal, perhaps incomplete, orbit segment whose length does not exceed L_k. Since both n_{k-1} and L_k are negligible in comparison with n_k (and hence with n), the two extreme pieces can be ignored and we get that the empirical measure associated with $\mathcal{S}|_{[1,n]\cap D}$ is (nearly) $2\varepsilon_{k-1}$-close to μ. Since k tends to infinity as n grows, $\mathcal{S}$ is generic for μ. □

LEMMA 3.6. *Let (X, T) be a topological dynamical system and let μ be an invariant measure on X. For each $\varepsilon > 0$ there exists $\delta > 0$ which satisfies the following assertion.*

Let $\mathcal{P}$ be a partition of X all of whose atoms have diameter not exceeding δ. Let $\mathcal{S}$ be a finite specification consisting of N_1 orbit segments of length l separated by some gaps (N_1 and l are arbitrary natural numbers, and the gaps are also arbitrary):

$$\mathcal{S}[a_N, a_N + l - 1] = x_N[0, l - 1],$$

where $a_1 \geq 0$ and, for each $N \in [1, N_1]$, we have $x_N \in X$ and $a_{N+1} - a_N \geq l$. Assume that for each $B \in \mathcal{P}^l = \bigvee_{n=0}^{l-1} T^{-n}(\mathcal{P})$ the frequency relative to $(a_N)_{N\in[1,N_1]}$,

$$\frac{|\{N \in [1, N_1] : x_N \in B\}|}{N_1},$$

is $\delta/|\mathcal{P}^l|$-close to $\mu(B)$ (this imposes that N_1 must in fact be huge). Then the empirical measure associated with $\mathcal{S}$,

$$\mu_{\mathcal{S}} = \frac{1}{|D|} \sum_{n\in D} \delta_{\mathcal{S}(n)},$$

where $D = \bigcup_{N=1}^{N_1} [a_N, a_N + l - 1]$ is the domain of $\mathcal{S}$, is ε-close to μ.

Proof. Regardless of what metric d compatible with the weak* topology on $\mathcal{M}(X)$ we are using, there exist a finite family of continuous [0, 1]-valued functions (say, $f_1, \ldots, f_K$) and a small positive number γ such that if

$$\left| \int f_k \, d\mu_1 - \int f_k \, d\mu_2 \right| < 3\gamma$$

for each $k \in [1, K]$, then $d(\mu_1, \mu_2) < \varepsilon$. Further, there exists β such that each of the finitely functions f_k varies on each β-ball in X by less than γ. We let $\delta = \min\{\beta, \gamma\}$. Let $\mathcal{P}$ be a partition of X as in the formulation of the lemma. Observe that if we replace each of the functions f_k by a function $\bar{f}_k$ constant on the atoms of $\mathcal{P}$ (say, assuming on each atom the supremum of f_k over that atom), then the integral of $\bar{f}_k$ with respect to any probability measure differs from the integral of f_k by at most γ. So, in order to show that $d(\mu_{\mathcal{S}}, \mu) < \varepsilon$, it suffices to show that

$$\left| \int f \, d\mu_{\mathcal{S}} - \int f \, d\mu \right| < \gamma,$$

for any [0, 1]-valued (not necessarily continuous) function f constant on the atoms of $\mathcal{P}$. For such a function f we have

$$\int f\, d\mu_{\mathcal{S}} = \frac{1}{|D|}\sum_{N=1}^{N_1}\sum_{n=0}^{l-1} f(T^n x_N) = \frac{1}{N_1}\sum_{N=1}^{N_1}\frac{1}{l}\sum_{n=0}^{l-1} f(T^n x_N)$$

(we are using the obvious fact that $|D| = N_1 l$). Note that if, for some $N, N' \in [1, N_1]$, the points x_N and $x_{N'}$ belong to the same atom B of $\mathcal{P}^l$ then the averages $(1/l)\sum_{n=0}^{l-1} f(T^n x_N)$ and $(1/l)\sum_{n=0}^{l-1} f(T^n x_{N'})$ are equal, so, we can replace them by $(1/l)\sum_{n=0}^{l-1} f(T^n x_B)$, where x_B is a point in B not depending on N. Then our integral becomes

$$\int f\, d\mu_{\mathcal{S}} = \sum_{B\in\mathcal{P}^l} \frac{|\{N \in [1, N_1] : x_N \in B\}|}{N_1}\frac{1}{l}\sum_{n=0}^{l-1} f(T^n x_B).$$

By assumption, the coefficient $|\{N \in [1, N_1] : x_N \in B\}|/N_1$ equals $\mu(B)$ up to $\delta/|\mathcal{P}^l|$, all the more so up to $\gamma/|\mathcal{P}^l|$. Since the averages $(1/l)\sum_{n=0}^{l-1} f(T^n x_B)$ do not exceed one, we obtain that $\int f\, d\mu_{\mathcal{S}}$ equals

$$\sum_{B\in\mathcal{P}^l} \mu(B)\frac{1}{l}\sum_{n=0}^{l-1} f(T^n x_B)$$

up to γ. Finally, observe that the latter sum equals $\int (1/l)\sum_{n=0}^{l-1} f\circ T^n\, d\mu$, which, by invariance of μ, equals $\int f\, d\mu$. We have shown that $|\int f\, d\mu_{\mathcal{S}} - \int f\, d\mu| < \gamma$, as required. □

The next proposition is our last preparatory fact before the proof of Theorem 1.3 It is also a standard fact (this time from ergodic theory), whose exact formulation is hard to find. Thus, we provide it with a proof.

PROPOSITION 3.7. *Let (X, T) be a topological dynamical system. Let x be a point quasi-generic for an ergodic measure μ and let $\mathcal{J} = (n_k)_{k\geq 1}$ be a sequence along which x generates μ. Fix a positive integer L. Then there exist two increasing sequences of positive integers, $(a_N)_{N\geq 0}$ and $(N_k)_{k\geq 1}$, satisfying the following conditions.*

(1) *For each $N \geq 1$ the difference $a_{N+1} - a_N$ equals either L or $L + 1$.*
(2) $\lim_{k\to\infty}(a_{N_k}/n_k) = 1$ *(that is, the sequences $(n_k)_{k\geq 1}$ and $(a_{N_k})_{k\geq 1}$ are equivalent).*
(3) *x generates μ with respect to the sequence $(a_N)_{N\geq 1}$, along $(N_k)_{k\geq 1}$, that is,*

$$\lim_{k\geq 1}\frac{1}{N_k}\sum_{N=1}^{N_k}\delta_{T^{a_N}(x)} = d\mu.$$

Proof. There exists an ergodic measure-preserving system (Y, ν, S) disjoint from (X, μ, T) (in the sense of Furstenberg). (Two measure-preserving systems are *disjoint* if their only joining is their product. An example of a system disjoint from (X, μ, T) is an irrational rotation by $e^{2\pi i t}$, where t is rationally independent from all numbers s such that $e^{2\pi i s}$ is an eigenvalue of (X, μ, T) (there are at most countably many values to be avoided).) By a standard application of Rokhlin towers, there exists a set A visited by ν-almost every orbit in Y infinitely many times with only two gap sizes between consecutive visits, L and $L+1$. There exists a topological model of (Y, ν, S) in which the set A is clopen, that is, its indicator function, denoted by F, is continuous. Let $y \in Y$ be a point generic for ν and let $(a_N)_{N\geq 1}$ denote the sequence of times of visits of the orbit of y in A (this sequence has only two gap sizes, L and $L+1$, as required in (1)). The pair (x, y) quasi-generates, along the sequence $\mathcal{J}$, some joinings of μ and ν. By disjointness, all such joinings are equal to the product measure $\mu \times \nu$ on $X \times Y$, that is, along $\mathcal{J}$ the pair (x, y) generates $\mu \times \nu$. This implies that, for any continuous function f on X,

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{n=1}^{n_k} f(T^n x)F(S^n y) = \int f\, d\mu \cdot \nu(A),$$

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{n=1}^{n_k} F(S^n y) = \nu(A).$$

Given $k \geq 1$, let N_k denote the largest N such that $a_N \leq n_k$. Observe that since $(a_N)_{N\geq 1}$ has bounded gaps, while $(n_k)_{k\geq 1}$ tends to infinity, the ratios a_{N_k}/n_k tend to one, as required in (2). Since $F(S^n y) = 1$ if and only if $n = a_N$ for some N (otherwise $F(S^n y) = 0$), we can rewrite the above limits as

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{N=1}^{N_k} f(T^{a_N} x) = \int f\, d\mu \cdot \nu(A), \tag{3.5}$$

$$\lim_{k\to\infty} \frac{N_k}{n_k} = \nu(A). \tag{3.6}$$

Dividing the left/right hand side of (3.5) by the left/right hand side of (3.6), we get

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{N=1}^{N_k} f(T^{a_N} x) = \int f\, d\mu.$$

Since this is true for any continuous function f on X, we have proved (3). □

4. *The main proof*

Proof of Theorem 1.3. Let $\mathcal{J} = (n_k)_{k\geq 1}$ be a sequence along which y generates ν. It suffices to construct a point x_0 such that the pair (x_0, y) generates ξ along a subsequence $\mathcal{J}'$ of $\mathcal{J}$. Clearly, such an x_0 generates μ along $\mathcal{J}'$ and, by Lemma 3.5, there will then exist a point x generic for μ and such that

$$\lim_{n\in\mathbb{M}} d(T^n x, T^n x_0) = 0,$$

where $\mathbb{M}$ is a set of upper density one achieved along a subsequence $\mathcal{J}''$ of $\mathcal{J}'$. Note that then the pair (x, y) still generates ξ along $\mathcal{J}''$, so the proof will be completed.

We fix a summable sequence of positive numbers $(\varepsilon_k)_{k\geq 1}$ and an increasing sequence of natural numbers l_k such that

$$\lim_{k\to\infty} \frac{M_{\varepsilon_k}(l_k)}{l_k} = 0. \tag{4.1}$$

Next, we let $(\mathcal{P}_k)_{k\geq 1}$ be a sequence of measurable partitions of X such that, for each $k \geq 1$, the diameters of the atoms of $\mathcal{P}_k$ do not exceed the number δ_k obtained from Lemma 3.6 for the measure μ and ε_k in the role of ε.

The atoms of the partitions $\mathcal{P}_k^l = \bigvee_{i=0}^{l-1} T^{-i}(\mathcal{P}_k)$, where $k \geq 1$ and $l \geq 1$, will be called *blocks* of X, while the atoms $\mathcal{P}_k^{l_k}$ will be called *blocks of order k of X*.

Likewise, we let $(\mathcal{Q}_k)_{k\geq 1}$ be a sequence of partitions of Y with diameters bounded by δ_k. We can easily arrange the partitions $\mathcal{Q}_k$ so that the orbit of y avoids the boundaries of the atoms of $\mathcal{Q}_k$ for each $k \geq 1$. (The partition $\mathcal{Q}_k$ can be constructed as follows. First we choose a finite open cover by δ_k-balls $B(c_i, \delta_k)$ centered at some points $c_i \in Y$, $i = 1, 2, \ldots, N$, $N \in \mathbb{N}$. There exists $\delta_k' < \delta_k$ such that for any $\delta \in [\delta_k', \delta_k]$, the balls $B(c_i, \delta)$ still cover Y. Note that, for each i, the δ-spheres $S(c_i, \delta)$ are disjoint for different values of δ. Since there are uncountably many values of δ while the orbit of y is countable, there exists a $\delta \in [\delta_k', \delta_k]$ such that all the spheres $S(c_i, \delta)$, $i = 1, 2, \ldots, N$, avoid the orbit of y. The partition $\mathcal{Q}_k$ is obtained by 'disjointification' of the cover by the balls $B(c_i, \delta)$. Then the boundaries of the atoms of $\mathcal{Q}_k$ are contained in the union of the δ-spheres $S(c_i, \delta)$, and hence the partition has the desired property.) The atoms of $\mathcal{Q}_k^l = \bigvee_{i=0}^{l-1} S^{-i}(\mathcal{Q}_k)$, where $k \geq 1$ and $l \geq 1$, will be called *blocks* of Y and the atoms of $\mathcal{Q}_k^{l_k}$ will be called *atoms of order k of Y*.

Note that if we apply the maximum metric in $X \times Y$ then the rectangular atoms of the partitions $\mathcal{P}_k \otimes \mathcal{Q}_k$ have diameters bounded by δ_k as well. We now choose some very small positive numbers γ_k so that, for each $k \geq 1$, we have

$$2\gamma_k + \gamma_k^2 < \frac{\delta_k}{|\mathcal{P}_k^l \otimes \mathcal{Q}_k^l|}.$$

Because y is generic for ν along $\mathcal{J}$, and its orbit avoids the boundaries of the blocks, the orbit of y visits each block C of Y with frequency evaluated at times n_k converging to $\nu(C)$.

Successively using Proposition 3.7 with the parameters $L_k = l_k + M_{\varepsilon_k}(l_k)$ in the role of L, and replacing, if necessary, the sequence $\mathcal{J}$ by a rapidly growing subsequence $\mathcal{J}'$ (from now on $(n_k)_{k\geq 1}$ will denote $\mathcal{J}'$), we can arrange two increasing sequences of positive integers, $(a_N)_{N\geq 1}$ and $(N_k)_{k\geq 0}$ starting with $N_0 = 0$, satisfying the following conditions.

(1) $\lim_{k\to\infty}(a_{N_k}/n_k) = 1$.

(2) For each $k \geq 1$ and each $N \in [N_{k-1}, N_k - 1]$, the difference $a_{N+1} - a_N$ equals either L_k or $L_k + 1$. (The proposition, as it is stated, does not allow us to ensure that the gap $a_{N_k} - a_{N_k - 1}$ (the first gap following the series of gaps of sizes L_k or $L_k + 1$) equals either L_{k+1} or $L_{k+1} + 1$. *A priori*, this gap may come out smaller (say, of size $j < L_{k+1}$). However, in this case, replacing the set A in the proof of the proposition (for $L = L_{k+1}$) by $T^{-(L_{k+1}-j)}A$, we can adjust the gap to size L_{k+1}.)

(3) If we denote by C_N the unique block of order k of Y containing $S^{a_N}y$, then, for any block C of order k of Y, we have

$$\left|\frac{1}{N_k - N_{k-1}}|\{N \in [N_{k-1}, N_k - 1] : C_N = C\}| - \nu(C)\right| < \gamma_k.$$

(4) If, in addition, C satisfies $\nu(C) > 0$, then also

$$|\{N \in [N_{k-1}, N_k - 1] : C_N = C\}| > \frac{1}{\gamma_k}.$$

Condition (1) says that $\mathcal{J}'$ and $\tilde{\mathcal{J}}' = (a_{N_k})_{k\geq 1}$ are equivalent. In particular, y generates ν along the sequence $(a_{N_k})_{k\geq 1}$ and if we find x_0 using $\tilde{\mathcal{J}}'$, the same x_0 will serve for $\mathcal{J}'$.

(*) Fix some $k \geq 1$. Let $\xi(B|C) = \xi(B \times C)/\nu(C)$, where B and C are blocks of order k of X and Y, respectively, with $\nu(C) > 0$. For every such C, the numbers $\xi(B|C)$, with B ranging over all blocks of order k of X, form a probability vector. By (4), this vector can be approximated up to γ_k (at each coordinate) by a rational probability vector with entries

$$\frac{r(B, C)}{|\{N \in [N_{k-1}, N_k - 1] : C_N = C\}|},$$

where each $r(B, C)$ is a non-negative integer. We can thus create a finite sequence $(B_N)_{N\in[N_{k-1},N_k-1]}$ of blocks of order k of X, so that, for every pair of blocks B, C of order k in X and Y respectively, we have

$$|\{N \in [N_{k-1}, N_k - 1] : C_N = C \text{ and } B_N = B\}| = r(B, C).$$

Then, for each pair B, C as above, with $\nu(C) > 0$, we have

$$\frac{r(B, C)}{N_k - N_{k-1}} = \frac{r(B, C)}{|\{N \in [N_{k-1}, N_k - 1] : C_N = C\}|} \cdot \frac{|\{N \in [N_{k-1}, N_k - 1] : C_N = C\}|}{N_k - N_{k-1}},$$

where (by the choice of the integers $r(B, C)$) the first fraction equals $\xi(B|C)$ up to γ_k, and, by (3), the second fraction equals $\nu(C)$, also up to γ_k. So, $r(B, C)/(N_k - N_{k-1})$ equals $\xi(B \times C)$ up to $2\gamma_k + \gamma_k^2$, which is less than $\delta_k/|\mathcal{P}_k^l \otimes \mathcal{Q}_k^l|$.

Now, we create a finite specification $\bar{\mathcal{S}}_k$ in $X \times Y$, as follows. For each $N \in [N_{k-1}, N_k - 1]$ we choose a point $x_N \in B_N$ and we let

$$\bar{\mathcal{S}}_k[a_N, a_N + L_k - 1] = (x_N, S^{a_N} y)[0, L_k - 1]$$

(the starting point of the Nth orbit segment falls in (B_N, C_N), and the second coordinate agrees, along the entire specification, with the orbit of y). Note that by (2), the gaps in the domain of $\bar{\mathcal{S}}_k$ have only two sizes, zero or one. Lemma 3.6 now guarantees that the empirical measure $\mu_{\bar{\mathcal{S}}_k}$ is ε_k-close to ξ.

Let $\bar{\mathcal{S}}$ be the infinite specification in $X \times Y$ defined as follows: for each $k \geq 1$ and each $N \in [N_{k-1}, N_k - 1]$ we let

$$\bar{\mathcal{S}}[a_N + M_{\varepsilon_k}(l_k), a_N + L_k - 1] = \bar{\mathcal{S}}_k[a_N + M_{\varepsilon_k}(l_k), a_N + L_k - 1].$$

It is fairly obvious that $\bar{\mathcal{S}}$ generates ξ along the sequence $\tilde{\mathcal{J}}'$ (by (4.1), the fact that the intervals of the domain are slightly trimmed on the left does not affect the convergence).

Let us denote by $\mathcal{S}$ the projection of $\bar{\mathcal{S}}$ to the first coordinate. This infinite specification in X satisfies all requirements of Lemma 3.1. That lemma allows us to find a point x_0 whose orbit shadows the specification $\mathcal{S}$ with an increasing accuracy. Clearly, the pair (x_0, y) shadows $\bar{\mathcal{S}}$ equally well. It is also clear that the domain of $\bar{\mathcal{S}}$ has density one, which (by Remark 3.2) implies that the pair (x_0, y) generates ξ along $\tilde{\mathcal{J}}'$, and hence also along $\mathcal{J}'$. We have achieved all that was necessary to complete the proof. □

Remark 4.1. It is possible to modify the proof and avoid the use of Lemma 3.5 (see below). Although the main proof itself becomes slightly longer, one can skip that lemma and the two auxiliary propositions altogether. There are two reasons why we have decided to present the longer argument.

(1) Lemma 3.5 is a generalization of Kamae's Theorem 1 in [**K**] and has some independent value of its own. It may turn out useful in further studies of systems with weak specification.
(2) By following the framework of the original proof, we show that T. Kamae has insightfully laid ground for further generalizations.

Sketch of the modified proof. Go to the paragraph marked by (*). Divide the blocks B (of order k of X) into two families: $\mathcal{B}$, of those whose associated empirical measures are close to μ, and the rest. By the mean ergodic theorem, for large enough k, the joint measure of the blocks in $\mathcal{B}$ is very close to one. Thus, by an insignificant renormalization, we can make the vector of conditional probabilities $\xi(B|C)$, with B ranging over $\mathcal{B}$ (and C fixed) probabilistic. From here we proceed as it is described except that each time we refer to B we use only the blocks from $\mathcal{B}$. The specification $\bar{\mathcal{S}}_k$ will then have its X-coordinate consisting exclusively of blocks $B \in \mathcal{B}$. The specification $\mathcal{S}$ (the X-projection of $\bar{\mathcal{S}}$) will consist of blocks whose empirical measures are getting closer and closer to μ. If the numbers a_{N_k} grow sufficiently fast in comparison to the lengths l_k then, by an identical argument to that in the proof of Lemma 3.5, the specification $\mathcal{S}$ will be generic for μ and so will be the point x_0 shadowing $\mathcal{S}$. Lemma 3.5 becomes irrelevant. □

Question 4.2. If y is generic for ν, can the pair (x, y) be obtained generic for ξ (as in the original theorem of Kamae)? Is a stronger specification property of X necessary for that?

Added in proof. While checking the proofs of our paper we became aware of an old paper of T. Kamae [**K1**] which contains results which partially overlap with ours. He introduces the vague separation property (v.s.p.) under which he proves a lifting theorem similar to ours. The relation between the v.s.p. and weak specification remains to be clarified.

REFERENCES

[BD] V. Bergelson and T. Downarowicz. On preservation of normality and determinism under arithmetic operations. In preparation.
[D] M. Dateyama. The almost weak specification property for ergodic group automorphisms of abelian groups. *J. Math. Soc. Japan* **42** (1990), 341–351.
[F] H. Furstenberg. Strict ergodicity and transformation of the torus. *Amer. J. Math.* **83** (1961), 573–601.
[FW] H. Furstenberg and B. Weiss. On almost 1-1 extensions. *Israel J. Math.* **65** (1989), 311–322.
[K] T. Kamae. Subsequences of normal sequences. *Israel J. Math.* **16** (1973), 121–149.
[K1] T. Kamae. Normal numbers and ergodic theory. *Proc. Third Japan–USSR Symp. on Probability Theory (Lecture Notes in Mathematics, 550)*. Ed. G. Maruyama and J. V. Prokhorov. Springer, Berlin, 1976.
[R] G. Rauzy. Nombres normaux et processus déterministes. *Acta Arith.* **29** (1976), 211–225.
[W] B. Weiss. Normal sequences as collectives. *Proc. Symp. on Topological Dynamics and Ergodic Theory*. University of Kentucky, Lexington, KY, 1971, pp. 79–80.