Finitary Reconstruction of a Measure Preserving Tranformation

Terrence M. Adams and Andrew B. Nobel *

April 23, 2000

Abstract

This paper considers the *finitary* reconstruction of an ergodic measure preserving transformation T of a complete separable metric space X from a single trajectory $x, Tx, \ldots$, or more generally, from a suitable reconstruction sequence $\mathbf{x} = x_1, x_2, \ldots$ with $x_i \in X$. An n-sample reconstruction is a function $T_n : X^{n+1} \to X$; the map $\hat{T}_n(\cdot; x_1, \ldots, x_n)$ is treated as an estimate of $T(\cdot)$ based on the n initial elements of $\mathbf{x}$. Given a reference probability measure μ_0 and constant $M > 1$, functions T_1, T_2, ... are defined, and it is shown that for every μ with $1/M \leq d\mu/d\mu_0 \leq M$, every μ-preserving transformation T, and every reconstruction sequence $\mathbf{x}$ for T, the estimates $\hat{T}_n(\cdot; x_1, \ldots, x_n)$ converge to T in the weak topology.

For the family of interval exchange transformations of $[0, 1)$ a simple family of estimates is described and shown to be consistent both pointwise and in the strong topology. However, it is also shown that no finitary estimation scheme is consistent in the strong topology for the family of all ergodic Lebesgue measure preserving transformations of the unit interval, even if $\mathbf{x}$ is assumed to be a generic trajectory of T.

Appears in *Israel Journal of Mathematics*, vol. 126, pp.309-326, 2001.

*T.M. Adams is with the Department of Mathematics and Computer Science, Rhode Island College. A.B. Nobel is with the Department of Statistics, University of North Carolina, Chapel Hill, NC 27514. Email: nobel@stat.unc.edu His work was supported in part by NSF Grant DMS-9501926.

1 Introduction

Let X be a complete separable metric space with Borel subsets $\mathcal{B}$ and probability measure μ, and suppose T is an ergodic measure preserving transformation (e.m.p.t.) of $(X, \mathcal{B}, \mu)$. It is well known that for μ-almost every $x \in X$, the (right infinite) trajectory $x, Tx, T^2x, \ldots$ of T starting at x determines T, where "determines" means that from knowledge of the *entire* trajectory one may, in principle, construct a map T' such that $\mu\{T' \neq T\} = 0$. Maharam [13] undertook a systematic study of individual sequences that determine measure preserving transformations of $[0, 1]$ in this infinitary sense. Her work was later extended by Bick [2], Kappos and Papadopoulou [9], Coffey [7], and Sun [19]. This paper considers the *finitary* reconstruction of an e.m.p.t. $T : X \to X$ from a single trajectory, or more generally, from an individual sequence whose successive entries determine the action of T. By finitary it is meant that successive estimates of T are produced from a given sequence in such a way that the n'th estimate depends only on the first n terms in the sequence.

Finitary reconstruction of Bernoulli processes has previously been considered by Ornstein and Weiss [17]. They describe an estimation scheme that, given any Bernoulli process $\mathbf{Y} = Y_1, Y_2, \ldots$, produces a sequence of processes $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ such that $\mathbf{Z}_k$ is constructed only from knowledge $Y_1, \ldots, Y_k$, and $\mathbf{Z}_k$ converges in the $\overline{d}$ distance to $\mathbf{Y}$. Moreover, they showed that no estimation scheme is $\overline{d}$ consistent for the larger family of K automorphisms. We consider here a weaker form of convergence, namely that of the weak topology, and in this setting it it is possible to define consistent estimates for a larger class of transformations.

To illustrate the notion of finitary reconstruction studied below, suppose that $T : X \to X$ is an ergodic, μ-preserving transformation of X. Recall that a sequence $S_1, S_2, \ldots$ of transformations is said to converge to T in the weak topology if $\mu(S_n^{-1} A \,\Delta\, T^{-1} A) \to 0$ for every $A \in \mathcal{B}$. Given a trajectory $x, Tx, T^2x, \ldots$ of T we wish to construct transformations $\hat{T}_1, \hat{T}_2, \ldots$ with the property that $\hat{T}_n$ depends only on $x, Tx, \ldots, T^{n-1}x$, and $\hat{T}_n \to T$ in the weak topology as $n \to \infty$. To see how such estimates might be constructed in a simple case, note that each pair $(T^{i-1}x, T^i x)$ of successive points in the trajectory lies on the graph of T. If $X = \mathbb{R}$ and T is piecewise continuous, then good estimates of T can easily be obtained by connecting neighboring points on its graph by straight lines; in higher dimensions linear interpolation will also suffice. (Section 5.1 shows how this can be done in the special case of infinite interval exchange transformations of $[0, 1]$.) Of interest here is the estimation of transformations T assumed only to be measurable, and in this case such simple estimates are not effective.

Finitary estimation is the sort most often considered in statistics, where it is commonly assumed that successive observations are independent, or obey one of a variety of mixing conditions. These mixing conditions give information about the rate at which sample averages converge to expectations. The principal difficulty encountered in the present context is the lack of rates of convergence in the ergodic theorem. As shown below, one may circumvent this difficulty if the measure preserved by T is comparable to a known reference measure.

1.1 Statement of Principal Result

If A is a subset of X then A^o, $\overline{A}$, and ∂A denote, respectively, the interior, closure, and boundary of A.

Definition: A sequence $\mathbf{x} = x_1, x_2, \ldots \in X$ is a *reconstruction sequence* for a measure preserving transformation T of $(X, \mathcal{B}, \mu)$ if

$$\frac{1}{n}\sum_{i=1}^{n} I\{x_i \in U\} I\{x_{i+1} \in V\} \to \mu(U \cap T^{-1}V) \tag{1}$$

for every $U, V \in \mathcal{B}$ such that $\mu(\partial U) = \mu(\partial V) = 0$. Note that the transformation T need not be ergodic. It is not assumed that the convergence in (1) takes place at any prespecified or uniform rate.

The following proposition follows in a routine manner from pointwise ergodic theorem.

Proposition 1 *Almost every trajectory of an e.m.p.t. $T : X \to X$ is a reconstruction sequence for T.*

A finitary *reconstruction scheme* is a sequence of measurable maps $T_n : X^{n+1} \to X$, $n \geq 1$. Given a reconstruction sequence $\mathbf{x} = x_1, x_2, \ldots$ for a μ-preserving transformation $T : X \to X$, the map

$$\hat{T}_n(x) = T_n(x; x_1, \ldots, x_n)$$

is taken to be an estimate of T based on the first n terms of $\mathbf{x}$. The scheme $\{T_n\}$ is *weakly consistent* for $\mathbf{x}$ if $\hat{T}_n \to T$ in the weak topology. Our goal is to exhibit a scheme that is simultaneously weakly consistent for a large family of transformations and reconstruction sequences. Let μ_0 be a non-atomic, reference probability measure on $(X, \mathcal{B})$, let $M > 1$, and define

$$\mathcal{D}(\mu_0, M) = \left\{ \mu : \frac{1}{M} \leq \frac{d\mu}{d\mu_0} \leq M \right\}.$$

Every measure $\mu \in \mathcal{D}(\mu_0, M)$ is finite and equivalent to μ_0. Our principal result is the following.

Theorem 1 *Given μ_0 and M there exists a finitary reconstruction scheme $\{T_n\}_{n=1}^{\infty}$ such that for each $\mu \in \mathcal{D}(\mu_0, M)$, each measure preserving transformation T of $(X, \mathcal{B}, \mu)$, and each reconstruction sequence $\mathbf{x} = x_1, x_2, \ldots$ for T, the estimates $\hat{T}_n(x) = T_n(x; x_1, \ldots, x_n)$ satisfy $\mu(\hat{T}_n^{-1} A \,\Delta\, T^{-1} A) \to 0$ for every $A \in \mathcal{B}$.*

The construction of the scheme $\{T_n\}_{n=1}^{\infty}$ is described in the proof of Theorem 1. To illustrate the result, suppose $X = [0, 1]$, μ_0 is Lebesgue measure, and $M = 1$, and let $T_n : [0, 1)^{n+1} \to [0, 1)$, $n \geq 1$, be the estimates of the theorem. Then for every ergodic Lebesgue measure preserving transformation T of $[0, 1)$, for almost every $x \in [0, 1)$, the maps $\hat{T}_n(\cdot) = T_n(\cdot; x, Tx, \ldots, T^{n-1}x)$ derived from the trajectory of T starting at x will converge to T in the weak topology.

The problem of estimating an iterated map has been considered in the context of chaos and non-linear dynamics, where the ultimate goal is typically prediction, or the estimation of some features of the dynamics such as Lyapunov exponents or the dimension of an attractor. Representative work can be found in the papers of Farmer and Sidorowich [8] , Casdagli [5, 6], Kostelich and Yorke [10], Nychka *et al.* [16], and Lu and Smith [12]. Additional work and references can be found in the book of Tong [20]. This work differs from that in the present paper in several respects. The cited references consider continuous or, more commonly, differentiable transformations, and their ultimate goal is to develop methods that are readily applicable to the analysis of experimental data. In addition, it is assumed that successive iterates of the transformation are perturbed by independent observational or dynamical noise, so that the actual trajectory of T is not directly observed. While the assumption of noisy observations complicates the problem in some respects, it enables one to apply time series and Markov chain techniques that are not applicable in the general setting considered here.

Lalley [11] describes a general means of reconstructing the orbit of a smooth diffeomorphism F, acting on a hyperbolic attractor, when the iterates of F are corrupted by additive, independent, observation noise. Bosq and Guégan [4] study kernel estimates of continuous, uniformly mixing transformations in the noiseless setting. Nobel and Adams [15] proposed finitary estimates, similar to the interpolation estimates of Section 2, for e.m.p.t.'s $T : \mathbb{R}^d \to \mathbb{R}^d$. Their estimates are L_1 consistent, but may not converge in the weak topology.

1.2 Summary

The next section is devoted to the problem of interpolation. The results established there are used in the proof of Theorem 1. In the interpolation problem the goal is to estimate a bounded function $f : X \to \mathbb{R}$ from a sequence of pairs $(x_1, y_1), (x_2, y_2), \ldots \in X \times \mathbb{R}$ that lie on the graph of f. A finitary interpolation scheme is propose, and its L_1-consistency is established in Theorem 2. An alternative characterization of reconstruction sequences is briefly described in Section 3. Section 4 contains the proof of Theorem 1.

Section 5 is devoted to the problem of finitary reconstruction in the strong topology, that is, reconstruction schemes for which $\mu\{\hat{T}_n \neq T\} \to 0$. It is shown that such schemes exist for the family of infinite affine interval exchange transformation of $[0, 1)$, but that no strongly consistent scheme exists for the larger family of ergodic Lebesgue measure preserving transformations of $[0, 1)$.

2 Interpolation of Bounded Functions

Our goal is to estimate a bounded, measurable function $f : X \to \mathbb{R}$ from a sequence of pairs $(x_1, y_1), (x_2, y_2), \ldots \in X \times \mathbb{R}$ whose limit points lie on the graph of f. Of particular interest is the case where $x_i = T^i x$ is the i'th point in the trajectory of a given ergodic transformation T, and $y_i = f(x_i) = f(T^i x)$ is the corresponding value of f. In this case (x_i, y_i) are points on the graph of f selected according to a trajectory of T.

Definition: Let $f : X \to \mathbb{R}$ be bounded and measurable, and let μ be a probability measure on $(X, \mathcal{B})$. A sequence $(\mathbf{x}, \mathbf{y}) = (x_1, y_1), (x_2, y_2), \ldots \in X \times \mathbb{R}$ is a *μ-interpolation sequence* for f if

(a) There exists $K < \infty$ such that $|y_i| \leq K$ for each $i \geq 1$.

(b) $n^{-1} \sum_{i=1}^n I\{x_i \in A\} \to \mu(A)$ if $\mu(\partial A) = 0$.

(c) $n^{-1} \sum_{i=1}^n y_i I\{x_i \in A\} \to \int_A f d\mu$ if $\mu(\partial A) = 0$.

(d) $n^{-1} \sum_{i=1}^n y_i^2 \to \int f^2 d\mu$.

When condition (b) holds, the sequence $\mathbf{x}$ is said to have one-dimensional stationary distribution μ.

Proposition 2 *If $f : X \to \mathbb{R}$ is bounded and T is an e.m.p.t. of $(X, \mathcal{B}, \mu)$, then for μ-almost every $x \in X$, $(\mathbf{x}, \mathbf{y}) = (x, f(x)), (Tx, f(Tx)), \ldots$ is a μ-interpolation sequence for f.*

A finitary interpolation scheme is a sequence of measurable maps $\phi_n : X \times (X \times \mathbb{R})^n \to \mathbb{R}$, $n \geq 1$. Given a μ-interpolation sequence $(\mathbf{x}, \mathbf{y})$ for f the map

$$\hat{\phi}_n(x) \;=\; \phi_n(x \,:\, (x_i, y_i)_{i=1}^n)$$

acts as an estimate of f based on $(x_1, y_1), \ldots, (x_n, y_n)$. The scheme $\{\phi_n\}_{n=1}^\infty$ is L_1 consistent for $(\mathbf{x}, \mathbf{y})$ if $\int |\hat{\phi}_n - f| \, d\mu \to 0$.

The interpolation and reconstruction schemes defined below are based on finite partitions of X with shrinking cells. Let μ_0 be a fixed reference probability measure on $(X, \mathcal{B})$, and let $\{u_1, u_2, \ldots\}$ be a countable dense subset of X. Define $B(x, r) = \{u : d(x, u) < r\}$ to be the open ball of radius r centered at x. For each i let $\{r_{i,j} : j \geq 1\}$ be positive numbers tending to zero such that $\mu_0(\,\partial B(u_i, r_{i,j})\,) = 0$. Let $\pi_0^o = \{X, \emptyset\}$, and for $k \geq 1$ define partitions $\pi_k^o = \bigwedge_{1 \leq i,j \leq k} B(x_i, r_{i,j})$. If $\pi[x]$ denotes the unique cell of π containing x, then for each $x \in X$,

$$\operatorname{diam}(\pi_k^o[x]) \to 0 \quad \text{as} \quad k \to \infty, \tag{2}$$

where $\operatorname{diam}(A) = \sup_{u,v \in A} d(u, v)$ denote the diameter of $A \subseteq X$. It follows that the semi-ring $\bigcup_{k \geq 0} \pi_k^o$ generates $\mathcal{B}$. Let $\pi_0 = \pi_0^o$, and let π_k be any partition that results from adjoining those elements of π_k^o having μ_0-measure zero to a fixed element of π_k^o having positive measure. The partitions $\pi_0, \pi_1, \ldots$ have the following properties:

(1) $\mu_0(\partial A) = 0$ for each $A \in \bigcup_{k \geq 0} \pi_k$

(2) $\mu_0(A) > 0$ for each $A \in \bigcup_{k \geq 1} \pi_k$

(3) $\bigcup_{k \geq 0} \pi_k$ is equivalent mod 0 to a semi-ring generating $\mathcal{B}$.

2.1 The Interpolation Scheme

Let μ_0 be the reference probability measure above. Fix $M > 1$, define

$$\mathcal{D}'(\mu_0, M) = \left\{ \mu \,:\, \mu \sim \mu_0 \;\; \text{and} \;\; \int \frac{d\mu_0}{d\mu} \, d\mu_0 \leq M^2 \right\}$$

Let $(\mathbf{x}, \mathbf{y}) = (x_1, y_1), (x_2, y_2), \ldots \in X \times \mathbb{R}$ be a μ-interpolation sequence for an unknown bounded function f. Given the first n terms $(x_1, y_1), \ldots, (x_n, y_n)$ of $(\mathbf{x}, \mathbf{y})$ define histograms

$$\phi_{k,n}(x) = \frac{\sum_{i=1}^n y_i I\{x_i \in \pi_k[x]\}}{\sum_{i=1}^n I\{x_i \in \pi_k[x]\}} \tag{3}$$

by averaging the values y_i within the cells of π_k, $k \geq 0$. If no x_i lies in $\pi_k[x]$ set $\phi_{k,n}(x) = 0$. Set

$$\Delta_{k,n} = \left(\frac{1}{n} \sum_{i=1}^n (\phi_{k,n}(x_i) - y_i)^2 \right)^{1/2} \tag{4}$$

equal to the empirical L_2-loss of $\phi_{k,n}$.

The histograms $\{\phi_{k,n} : k \geq 1\}$ are candidate estimates of f. An estimate $\hat{\phi}_n$ is selected from among the candidates by adaptively choosing a suitable index k_n based on $\Delta_{k,n}$. Fix constants $\epsilon_1 > \epsilon_2 > \cdots > 0$ tending to zero. Let k_n be the largest integer $k \geq 1$ such that

$$\int |\phi_{l,n} - \phi_{j,n}| \, d\mu_0 \;\leq\; 2M\Delta_{j,n} + 2(1+M)\epsilon_j \quad \text{for} \quad 1 \leq j \leq l \leq k \tag{5}$$

and define

$$\hat{\phi}_n(x) \;=\; \hat{\phi}_n(x : x_1, y_1, \ldots, x_n, y_n) \;=\; \phi_{k_n,n}(x)\,. \tag{6}$$

In most cases the index k_n will not increase monotonically with the sample size n, nor will it grow at a prespecified rate.

Theorem 2 *For every measure $\mu \in \mathcal{D}'(\mu_0, M)$, every bounded measurable function $f : X \to$ $\mathbb{R}$, and every μ-interpolation sequence $(\mathbf{x}, \mathbf{y})$ for f, the estimates $\hat{\phi}_n$ produced according to (3)-(6) are such that $\int_X |\hat{\phi}_n - f| \, d\mu \to 0$ as $n \to \infty$.*

Remarks: The proposed estimates are finitary, as $\hat{\phi}_n$ depends on $(x_1, y_1), \ldots, (x_n, y_n)$. The theorem requires only that the one dimensional distribution μ of $\mathbf{x}$ be comparable to μ_0. Beyond condition (c) above, no assumptions are placed on the joint behavior of $\mathbf{y}$ and $\mathbf{x}$. Note that evaluation of $\hat{\phi}_n$ requires knowledge of both μ_0 and M.

2.2 Proof of Theorem 2

Let μ be a Borel probability measure on X and let $f : X \to \mathbb{R}$ be bounded and measurable. For each finite partition π of X let

$$(f \circ \pi)(x) = \frac{1}{\mu(\pi[x])} \int_{\pi[x]} f(v) d\mu(v),$$

provided that $\mu(\pi[x]) > 0$, and set $(f \circ \pi)(x) = 0$ otherwise. Thus $f \circ \pi$ is a version of the conditional expectation of f given π. Define the $L_2(\mu)$ norm $||f|| = \left(\int |f|^2 \, d\mu\right)^{1/2}$, and note that $||f \circ \pi|| \leq ||f||$.

Lemma 1 *Let $(\mathbf{x}, \mathbf{y}) = (x_1, y_1), (x_2, y_2), \ldots$ be a μ-interpolation sequence for a bounded function f and let $f_k = (f \circ \pi_k)$. If $\mu \sim \mu_0$ then*

(i) $||f_k - f||$ decreases to zero as $k \to \infty$

(ii) $\max_{x \in X} |\phi_{k,n}(x) - f_k(x)| \to 0$ for each $k \geq 1$

(iii) $\Delta_{k,n} \to ||f_k - f||$ *for each* $k \geq 1$.

Proof: As the partitions π_k are nested, part (i) of the lemma follows directly from the martingale convergence theorem and standard properties of conditional expectations. Properties (a) and (b) of $(\mathbf{x}, \mathbf{y})$ readily imply the pointwise convergence of $\phi_{k,n}$ to f_k. As each function is constant on the cells of the finite partition π_k, assertion (ii) of the lemma follows. To establish (iii), define $\tilde{\Delta}_{k,n}$ as in equation (4), with f_k in place of $\phi_{k,n}$. Then

$$| \Delta_{k,n} - ||f_k - f|| \, | \;\leq\; | \Delta_{k,n} - \tilde{\Delta}_{k,n} | \;+\; | \tilde{\Delta}_{k,n} - ||f_k - f|| \, | \,.$$

The first term on the right hand side is at most

$$\left(\frac{1}{n} \sum_{i=1}^{n} |f_k(x_i) - \phi_{k,n}(x_i)|^2 \right)^{1/2} \;\leq\; \max_{x \in X} |\phi_{k,n}(x) - f_k(x)| \,,$$

which tends to zero by (ii). As for the second term,

$$\tilde{\Delta}_{k,n}^2 \;=\; \frac{1}{n} \sum_{i=1}^{n} (f_k(x_i) - y_i)^2 \;=\; \frac{1}{n} \sum_{i=1}^{n} (\, f_k^2(x_i) - 2 f_k(x_i) y_i + y_i^2 \,) \,.$$

The average of y_i^2 converges to $\int f^2 d\mu$ by property (d) of $(\mathbf{x}, \mathbf{y})$. Let c_j be the value of f_k on the cell $A_j \in \pi_k$. As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} f_k^2(x_i) \;=\; \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{r_k} c_j^2 I_{A_j}(x_i) \;\to\; \sum_{j=1}^{r_k} c_j^2 \mu(A_j) \;=\; \int f_k^2 d\mu$$

by virtue of (a), and

$$\frac{1}{n} \sum_{i=1}^{n} f_k(x_i) f(x_i) \;=\; \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{r_k} c_j I_{A_j}(x_i) f(x_i) \;\to\; \sum_{j=1}^{r_k} c_j \int_{A_j} f \, d\mu \;=\; \int f_k f \, d\mu$$

by virtue of (b). Comparing these limits with the corresponding terms of $||f_k - f||^2$ completes the proof. $\Box$

Proof of Theorem 2: Let μ be a probability measure in $\mathcal{D}'(\mu_0, M)$, and let $(\mathbf{x}, \mathbf{y})$ be any μ-interpolation sequence for a bounded function $f : X \to \mathbb{R}$. Fix $k \geq 1$. By Lemma 1 there exists $N = N(k)$ so large that for each $n \geq N$ and each $1 \leq j \leq k$,

$$\max_{x \in X} |\phi_{j,n}(x) - f_j(x)| \;\leq\; \epsilon_j \;\text{ and }\; ||f_j - f|| \;\leq\; \Delta_{j,n} + \epsilon_j \tag{7}$$

(recall that $|| \cdot ||$ is the $L_2(\mu)$ norm). For each $n \geq N$ and each $1 \leq j \leq l \leq k$,

$$\int |\phi_{l,n} - \phi_{j,n}| \, d\mu_0 \;\;\leq\;\; \int |f_l - f_j| \, d\mu_0 \;+\; 2\epsilon_j$$

$$
\begin{aligned}
&\le \; ||f_l - f_j|| \cdot \left(\int (d\mu_0/d\mu)\, d\mu_0\right)^{1/2} + 2\epsilon_j \\
&\le \; M(\,||f_l - f|| + ||f_j - f||\,) + 2\epsilon_j \\
&\le \; 2M\,||f_j - f|| + 2\epsilon_j \\
&\le \; 2M\Delta_{j,n} + 2(1+M)\epsilon_j\,.
\end{aligned}
$$

where the second step is a consequence of the Cauchy-Schwartz inequality. It follows from (5) that $k_n \to \infty$.

Fix $r \ge 1$ and let $N' = N'(r)$ be so large that $k_n \ge r$, $\Delta_{r,n} \le ||f_r - f||_2 + \epsilon_r$, and $\max_{x\in X} |\phi_{r,n}(x) - f_r(x)| < \epsilon_r$ for each $n \ge N'$. When $n, m \ge N'$, inequality (5) implies that

$$
\begin{aligned}
\int |\hat\phi_n - \hat\phi_m|\, d\mu_0 &\le \int |\phi_{k_n,n} - \phi_{r,n}|\, d\mu_0 + \int |\phi_{k_m,m} - \phi_{r,m}|\, d\mu_0 + 2\epsilon_r \\
&\le 2M(\Delta_{r,m} + \Delta_{r,n}) + (6+4M)\epsilon_r \\
&\le 4M||f_r - f||_2 + 6(1+M)\epsilon_r\,. \qquad (8)
\end{aligned}
$$

Suitable choice of r makes the last sum less than any given positive number. Thus $\{\hat\phi_n\}$ is a Cauchy sequence in $L_1(d\mu_0)$, and there is therefore an integrable function f^* for which $\int |\hat\phi_n - f^*|\, d\mu_0 \to 0$. For each n such that $k_n \ge r$,

$$
\begin{aligned}
&\int |f - f^*|\, d\mu_0 \\
&\le \int |f - f_r|\, d\mu_0 + \int |f_r - \phi_{r,n}|\, d\mu_0 + \int |\phi_{r,n} - \phi_{k_n,n}|\, d\mu_0 + \int |\phi_{k_n,n} - f^*|\, d\mu_0 \\
&\le M||f - f_r|| + \int |f_r - \phi_{r,n}|\, d\mu_0 + 2M\Delta_{r,n} + 2(1+M)\epsilon_r + \int |\hat\phi_n - f^*|\, d\mu_0\,.
\end{aligned}
$$

Letting n, and then r, tend to infinity shows that $\int |f - f^*|\, d\mu_0 = 0$. As μ is dominated by μ_0 the proof is complete. □

3 Reconstruction and Predictive Sequences

Here we give an alternative characterization of reconstruction sequences. A sequence $\mathbf{x} = x_1, x_2, \ldots$ with $x_i \in X$ is stable if

$$
\Lambda(g) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i) \qquad (9)
$$

exists for every bounded continuous function $g : X \to \mathbb{R}$, and is predictive if for every $\epsilon > 0$ there is a compact set K and a continuous function $h : K \to X$ such that

$$
\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I\{x_i \notin K\} \;\le\; \epsilon \qquad (10)
$$

and

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I\{x_i \in K \text{ and } d(h(x_i), x_{i+1}) \geq \epsilon\} \;\leq\; \epsilon. \tag{11}$$

Condition (9) indicates that $\mathbf{x}$ has limiting first order relative frequencies, while conditions (10) and (11) ensure that the elements of $\mathbf{x}$ are concentrated on compact sets, and that on these sets one can predict x_{i+1} by a continuous function of x_i with small average error.

It is shown in Nobel [14] that if $\mathbf{x}$ is stable and predictive then there exists a unique measure μ on $(X, \mathcal{B})$ and a unique m.p.t. T of $(X, \mathcal{B}, \mu)$ such that $\mathbf{x}$ is a reconstruction sequence for T. Conversely, if X is a separable Banach space, then every reconstruction sequence $\mathbf{x}$ with values in X is stable and predictive.

4 Proof of Theorem 1

An explicit construction of the scheme $\{T_n\}_{n=1}^{\infty}$ in Theorem 1 is given below. The construction relies on the interpolation procedure of Theorem 2. In special cases, e.g. when $X = [0,1]$ and π_k is the k'th dyadic partition of the unit interval, the estimates $\hat{T}_n$ can be constructed by a computer, though not in an efficient fashion.

Proof: Let $\mathbf{x}$ be a reconstruction sequence for a μ-preserving transformation $T : X \to X$. Fix $k \geq 1$ for the moment and write $\pi_k = \{A(j,k) : 1 \leq j \leq s(k)\}$. For each $x \in X$ let $\pi_k(x)$ be the unique integer $j \in \{1, \ldots, s(k)\}$ such that $x \in A(j,k)$.

Claim 1: The sequence $(x_i, y_i) = (x_i, \pi_k(x_{i+1}))$, $i \geq 1$, is a μ-interpolation sequence for the function

$$g_k(x) = \sum_{j=1}^{s(k)} j I\{x \in T^{-1} A_{j,k}\}\,.$$

Proof: Clearly $|y_i| \leq s(k) < \infty$ for each $i \geq 1$. Condition (b) follows from (1) when $V = X$. Moreover, (1) implies that if $\mu_0(\partial A) = 0$ then as n tends to infinity,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} y_i I\{x_i \in A\} \quad &= \quad \sum_{j=1}^{s(k)} \frac{1}{n}\sum_{i=1}^{n} j I\{x_i \in A\} I\{x_{i+1} \in A(j,k)\} \\
&\to \quad \sum_{j=1}^{s(k)} j\mu(A \cap T^{-1}A(j,k)) \;=\; \int_A g_k \, d\mu\,,
\end{aligned}$$

which is (c). Condition (d) follows similarly as

$$\frac{1}{n}\sum_{i=1}^{n} y_i^2 \;=\; \sum_{j=1}^{s(k)} \frac{1}{n}\sum_{i=1}^{n} j^2 I\{x_{i+1} \in A(j,k)\} \;\to\; \sum_{j=1}^{s(k)} j^2 \mu(T^{-1}A(j,k)) \;=\; \int g_k^2 \, d\mu. \quad \Box$$

Now let $x_1, \ldots, x_n$ be the first n terms of $\mathbf{x}$, on the basis of which an estimate $\hat{T}_n$ of T will be created. For each $k \geq 1$ let

$$(x_1, \pi_k(x_2)), (x_2, \pi_k(x_3)), \ldots, (x_{n-1}, \pi_k(x_n)) \tag{12}$$

be pairs derived from the given finite sequence and π_k. Applying the interpolation procedure of Theorem 2 to these pairs yields an estimate $\hat{g}_{k,n}$ of g_k. Let

$$\tilde{g}_{k,n}(x) \;=\; \min\{\, 1 \leq j \leq s(k) : |\hat{g}_{k,n}(x) - j| \leq 1/2 \,\}$$

be a discretized version of this estimate, and define $B_n(j,k) \;=\; \tilde{g}_{k,n}^{-1}(j)$.

Claim 2: For large n the sets $B_n(j,k)$ approximate $T^{-1}A(j,k)$, in the sense that

$$\lim_{n\to\infty} \sum_{j=1}^{s(k)} \mu_0(B_n(j,k) \,\Delta\, T^{-1}A(j,k)) \;=\; 0\,.$$

Proof: Claim 1 and Theorem 2 jointly imply that $\int |\hat{g}_{k,n} - g_k| \, d\mu_0 \to 0$ as $n \to \infty$, and it follows that

$$\lim_{n\to\infty} \int |\tilde{g}_{k,n} - g_k| \, d\mu_0 \;=\; 0\,. \tag{13}$$

As each family $\{A(j,k) : 1 \leq j \leq s(k)\}$ and $\{B_n(j,k) : 1 \leq j \leq s(k)\}$ consists of disjoint sets,

$$\begin{aligned}
&\sum_{j=1}^{s(k)} \mu_0(B_n(j,k) \,\Delta\, T^{-1}A(j,k)) \\
&\quad = \;\; \sum_{j=1}^{s(k)} \int |\, I\{x \in B_n(j,k)\} - I\{x \in T^{-1}A(j,k)\} \,| \, d\mu_0 \\
&\quad \leq \;\; 2 \int |\, \sum_{j=1}^{s(k)} jI\{x \in B_n(j,k)\} - \sum_{j=1}^{s(k)} jI\{x \in T^{-1}A(j,k)\} \,| \, d\mu_0 \\
&\quad \leq \;\; 2 \int |\hat{g}_{k,n} - g_k| \, d\mu_0\,,
\end{aligned}$$

and the claim follows from (13). The stated convergence also holds for μ as $\mu \sim \mu_0$. $\Box$

Properties (1) and (2) of the partitions π_k ensure that $\mu_0(A(j,k)) > 0$ and $\mu_0(\partial A(j,k)) = 0$ for each $k \geq 1$ and $1 \leq j \leq s(k)$. By definition, $B_n(j,k) \;=\; \tilde{g}_{k,n}^{-1}(j)$ is a union of cells $A \in \pi_r$, where r is selected by the interpolation procedure of Theorem 2, and therefore $\mu_0(\partial B(j,k)) = 0$. Since μ_0 is non-atomic there exists for each n, j, k such that $\mu_0(B_n(j,k)) > 0$ a measurable map $\alpha_{n,j,k} : \overline{B}_n(j,k) \to \overline{A}(j,k)$ that preserves normalized μ_0-measure in the sense that

$$\frac{\mu_0(\alpha_{n,j,k}^{-1}C)}{\mu_0(B_n(j,k))} \;=\; \frac{\mu_0(\alpha_{n,j,k}^{-1}C)}{\mu_0(\overline{B}_n(j,k))} \;=\; \frac{\mu_0(C)}{\mu_0(\overline{A}(j,k))} \;=\; \frac{\mu_0(C)}{\mu_0(A(j,k))} \tag{14}$$

for each measurable set $C \subseteq A(j,k)$ (c.f. Royden [18], Theorem 15.5.16). If $\mu_0(B_n(j,k)) = 0$, let $\alpha_{n,j,k}$ map $\overline{B}_n(j,k)$ to a single point $x_0 \in X$. For each $k \geq 1$ define candidate estimates

$$T_{k,n}(x) \;=\; \sum_{j=1}^{s(k)} \alpha_{n,j,k}(x)\, I\{x \in B_n(j,k)\}. \tag{15}$$

Fix positive numbers $\epsilon_1 > \epsilon_2 > \cdots$ tending to zero, and let k_n be the greatest $k \geq 1$ such that

(a) $\sum_{A \in \pi_l} \mu_0(\, T_{l,n}^{-1} A \,\Delta\, T_{k,n}^{-1} A\,) \;\leq\; \epsilon_l$ for each $l \leq k$.

(b) $0 \;<\; \mu_0(B_n(j,k)) \;\leq\; 2M^2 \cdot \mu_0(A(j,k))$ for $1 \leq j \leq s(k)$.

Definition: The estimate $\hat{T}_n$ based on the initial sequence $x_1, \ldots, x_{n+1}$ of $\mathbf{x}$ is $T_{k_n,n}$.

Claim 3: The index k_n tends to infinity as n tends to infinity.

Proof: Fix $k \geq 1$ and let $l \leq k$ and $1 \leq j \leq s(l)$. Note that

$$\begin{aligned}
\mu_0(\, T_{l,n}^{-1} A(j,l) \;\Delta\; T_{k,n}^{-1} A(j,l)\,) \;&=\; \mu_0(\, B_n(j,l) \;\Delta\; T_{k,n}^{-1} A(j,l)\,) \\
\leq\;\; &\mu_0(\, B_n(j,l) \;\Delta\; T^{-1} A(j,l)\,) \;+\; \mu_0(\, T^{-1} A(j,l) \;\Delta\; T_{k,n}^{-1} A(j,l)\,)
\end{aligned}$$

Now $A(j,l)$ is equivalent mod zero to a union $\cup_{i=1}^m A_i$ of cells $A_i \in \pi_k$, so the second term above is equal to

$$\begin{aligned}
\mu_0(\, T^{-1} \cup_{i=1}^m A_i \;\Delta\; T_{k,n}^{-1} \cup_{i=1}^m A_i\,) \;\;&\leq\;\; \sum_{i=1}^{m} \mu_0(\, T^{-1} A_i \;\Delta\; T_{k,n}^{-1} A_i\,) \\
&\leq\;\; \sum_{i=1}^{s(k)} \mu_0(\, T^{-1} A(i,k) \;\Delta\; B_n(i,k)\,)
\end{aligned}$$

Thus for $l \leq k$ the sum in condition (a) above is at most

$$\sum_{i=1}^{s(l)} \mu_0(\, T^{-1} A(i,l) \;\Delta\; B_n(i,l)\,) \;+\; s(l) \cdot \sum_{i=1}^{s(k)} \mu_0(\, T^{-1} A(i,k) \;\Delta\; B_n(i,k)\,),$$

which tends to zero as $n \to \infty$ by Claim 2. Therefore (a) holds when n is sufficiently large. As $\mu \sim \mu_0$ and π_k is finite, Claim 2 also implies that as $n \to \infty$,

$$\begin{aligned}
\sum_{i=1}^{s(k)} |\mu(B_n(i,k)) - \mu(A(i,k))| \;\;&=\;\; \sum_{i=1}^{s(k)} |\mu(B_n(i,k)) - \mu(T^{-1} A(i,k))| \\
&\leq\;\; \sum_{i=1}^{s(k)} \mu(\, B_n(i,k) \,\Delta\, T^{-1} A(i,k)\,) \;\to\; 0\,.
\end{aligned} \tag{16}$$

As $\mu_0(A(j,k)) > 0$ for each $1 \leq j \leq s(k)$, the same is true of $\mu(A(j,k))$, and as $\mu \in \mathcal{D}(\mu_0, M)$,

$$\frac{\mu_0(B_n(j,k))}{\mu_0(A(j,k))} \;\leq\; M^2 \cdot \frac{\mu(B_n(j,k))}{\mu(A(j,k))}\,.$$

It then follows from condition (16) that (b) holds when n is sufficiently large. $\square$

Claim 4: For every $C \in \mathcal{B}$, $\lim_{n\to\infty} \mu_0(\hat{T}_n^{-1} C \,\Delta\, T^{-1} C) = 0$.

Proof: First consider $C \in \pi_s$ and let $l \geq s$. Then C is equivalent mod zero to a union of cells $A \in \pi_l$, and therefore

$$\mu_0(T^{-1} C \,\Delta\, \hat{T}_n^{-1} C) \;\;\leq\;\; \sum_{A \in \pi_l} \mu_0(T^{-1} A \,\Delta\, T_{l,n}^{-1} A) \;+\; \sum_{A \in \pi_l} \mu_0(T_{l,n}^{-1} A \,\Delta\, T_{k_n,n}^{-1} A)\,.$$

As k_n tends to infinity, Claim 2 implies that

$$\limsup_{n\to\infty} \mu_0(T^{-1} C \,\Delta\, \hat{T}_n^{-1} C) \;\leq\; \epsilon_l\,,$$

and the asserted convergence follows as l was arbitrary.

Now let C be any element of $\mathcal{B}$. Fix $\epsilon > 0$ and select $\delta > 0$ such that $\mu(B) < \delta$ implies $\mu_0(B) < \epsilon$. As $\cup_{k=0}^{\infty} \pi_k$ is equivalent mod zero to a semi-ring of sets generating $\mathcal{B}$, there exists $s \geq 1$ and sets $\{A_\alpha\} \subseteq \pi_s$ for which $E = C \,\Delta\, (\cup A_\alpha)$ satisfies $\mu(E) < \delta$. The value of $\mu_0(T^{-1} C \,\Delta\, \hat{T}_n^{-1} C)$ is at most

$$\mu_0(T^{-1} E) \;+\; \mu_0(T^{-1}(\cup A_\alpha) \,\Delta\, \hat{T}_n^{-1}(\cup A_\alpha)) \;\;+\; \mu_0(\hat{T}_n^{-1} E)\,.$$

As $\mu(T^{-1} E) \;=\; \mu(E) \;<\; \delta$ the first term above is less than ϵ. For each $l \geq s$ the second term is at most

$$\sum_{A \in \pi_l} \mu_0(T^{-1} A \,\Delta\, \hat{T}_n^{-1} A) \;\leq\; \sum_{A \in \pi_l} \mu_0(T^{-1} A \Delta T_{l,n}^{-1} A) \;+\; \sum_{A \in \pi_l} \mu_0(T_{l,n}^{-1} A \Delta T_{k_n,n}^{-1} A),$$

which tends to zero by arguments give in the special case $C \in \pi_s$ treated above. Condition (a) in the definition of k_n implies that

$$\begin{aligned}
\mu_0(\hat{T}_n^{-1} E) \;&=\; \int_E \frac{d(\mu_0 \circ \hat{T}_n^{-1})}{d\mu_0}\, d\mu_0 \\
&=\; \sum_{j=1}^{s(k_n)} \int_{E \cap A(j,k_n)} \frac{d(\mu_0 \circ \hat{T}_n^{-1})}{d\mu_0}\, d\mu_0 \\
&=\; \sum_{j=1}^{s(k_n)} \int_{E \cap A(j,k_n)} \frac{\mu_0(\hat{B}_n(j,k_n))}{\mu_0(A(j,k_n))}\, d\mu_0 \\
&\leq\; \sum_{j=1}^{s(k_n)} 2M^2 \cdot \mu_0(E \cap A(j,k_n)) \\
&=\; 2M^2 \mu_0(E) \;\leq\; 2M^2 \epsilon\,.
\end{aligned}$$

As $\epsilon > 0$ was arbitrary, the proof of the claim is complete. The assertion of Theorem 1 follows as $\mu_0 \sim \mu$. □

5 Strong Topology

The strong topology on the space of measurable transformations of $[0, 1)$ is the topology induced by the metric

$$d(S, T) = \lambda\{x : Sx \neq Tx\}.$$

where λ denotes Lebesgue measure on $[0, 1)$. In contrast with Theorem 1 above, it is shown in Theorem 3 that it is not possible to estimate every Lebesgue measure preserving transformation of $[0, 1)$ in the strong topology from finite segments of its orbit. To do this, a measure ν is placed on a family $\mathcal{S}$ of e.m.p.t.'s derived from the von Neumann-Kakutani adding machine. The richness of $\mathcal{S}$ under ν is then used to show that there is no consistent procedure for estimating transformations in $\mathcal{S}$ in the strong topology.

On the other hand, it is possible to obtain strongly consistent estimates for restricted families of transformations. To illustrate this, the next section exhibits a consistent scheme for estimating any infinite interval exchange transformation of $[0, 1)$ in the strong topology.

5.1 Infinite Interval Exchanges

Many examples of ergodic Lebesgue measure preserving transformations appearing in the literature are defined on the unit interval as infinite interval exchange transformations. In particular, every measure preserving transformation defined on a non-atomic, separable probability space is measure theoretically isomorphic to an infinite interval exchange transformation defined on $[0, 1)$ with Lebesgue measure (Arnoux, Ornstein and Weiss [1]). There is a simple procedure that provides consistent finitary estimates of every interval exchange map in the strong topology. The procedure is linear interpolation, accomplished by connecting adjacent points on the graph with straight line segments. This procedure will work for a wider class of maps defined on $[0, 1)$. Let I_j, $j \geq 1$, be disjoint subintervals of $[0, 1)$, and let α_j be real constants. A map f is an $E\{I_j, \alpha_j\}$-map if $\lambda(\cup_{j=1}^{\infty} I_j) = 1$ and $f'(x) = \alpha_j$ for $x \in I_j$. Given n sample pairs $(x_i, y_i) = (x_i, f(x_i))$ of such a map, order the x_i so that $x_1 < x_2 < \ldots < x_n$, and let $x_0 = 0$, $x_{n+1} = 1$, $y_0 = y_1$ and $y_{n+1} = y_n$. For $1 \leq i \leq n+1$, define

$$\hat{\alpha}_{n,i} = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$$

and let

$$\hat{f}_n(x) = y_i + \hat{\alpha}_{n,i}(x - x_i)$$

for $x_{i-1} \le x < x_i$.

Proposition 3 *If f is an $E\{I_j, \alpha_j\}$-map and $\{x_i : i = 1, 2, \ldots\} \subset \bigcup_{j=1}^{\infty} I_j$ is dense in $[0,1)$, then for almost every $x \in [0,1)$, $\hat{f}_n(t) = f(x)$ for all but a finite number of n.*

Proof: Suppose that x is contained in I_k for some k. Since the sequence x_i is dense, there exist positive integers n and $i \le n$ such that $a_k < x_{i-1} \le x < x_i < b_k$. Since f is linear on (a_k, b_k), $\hat{\alpha}_{n,i} = \alpha_k$. Therefore,

$$\hat{f}_n(x) = y_i + \hat{\alpha}_{n,i}(x - x_i) = f(x_i) + \alpha_k(x - x_i) = f(x)$$

and $\hat{f}_m(x) = f(x)$ for all $m \ge n$. $\square$

Corollary 1 *If T is an infinite interval exchange transformation then for almost every $x \in [0,1)$ the estimates $\hat{T}_n$ formed as above from pairs $(x_i, y_i) = (T^{i-1}x, T^i x)$ will converge to T pointwise and in the strong topology.*

5.2 A Counterexample

Theorem 3 *Let $\mathcal{T}$ be the family of Lebesgue measure preserving transformations of $[0,1)$. No reconstruction scheme $\{T_n\}$ has the property that $\lambda\{u : T_n(u : x, \ldots, T^{n-1}x) \ne T(u)\} \to 0$ for every $T \in \mathcal{T}$ and every $x \in [0,1]$ such that $x, Tx, \ldots$ is a reconstruction sequence for T.*

Proof: Given a stack $C = \{I_1, \ldots, I_{2m}\}$ of intervals $I_i = [a_i, b_i) \subseteq [0,1)$, define the switching map β_C on $\bigcup_{i=m+1}^{2m} I_i$ by

$$\beta_C(x) = \begin{cases} x + (\frac{b_i - a_i}{2}) & \text{if } x \in [a_i, \frac{a_i + b_i}{2}) \\ x - (\frac{b_i - a_i}{2}) & \text{if } x \in [\frac{a_i + b_i}{2}, b_i) \end{cases}$$

for $m + 1 \le i \le 2m$. Let T be the von Neumann-Kakutani adding machine and let $C_n = \{I_1, \ldots, I_{m_n}\}$, $n \ge 1$, be the columns formed in the construction of T where $m_n = 2^{n-1}$. Let $\beta_i = \beta_{C_{i+2}}$ for positive integers i. Define the set $\mathcal{S}$ of switching sequences

$$\mathcal{S} = \{(\phi_i)_{i=1}^{\infty} : \phi_i \in \{\beta_i, \text{identity}\}\}.$$

For each $\phi = (\phi_i) \in \mathcal{S}$, the map

$$T_\phi = \lim_{n \to \infty} (\phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_1 \circ T)$$

is an invertible, ergodic, Lebesgue measure preserving transformation of $[0,1)$. Note the distance between any two distinct elements ϕ and ψ in $\mathcal{S}$:

$$\lambda\{x : T_\phi x = T_\psi x\} \geq \frac{1}{2}.$$

Let ν be Bernoulli$(\frac{1}{2}, \frac{1}{2})$ measure on $\mathcal{S}$. For each positive integer q, let

$$\mathcal{S}_q = \{\phi = (\phi_1, \phi_2, \ldots) : \phi_q \text{ is the identity}\}.$$

Define the map $\xi_q : \mathcal{S} \to \mathcal{S}$ by

$$\xi_q(\phi_1, \ldots, \phi_{q-1}, \phi_q, \phi_{q+1}, \ldots) = (\phi_1, \ldots, \phi_{q-1}, \bar{\phi_q}, \phi_{q+1}, \ldots)$$

where $\bar{\phi_q} \neq \phi_q$. Thus the restriction of ξ_q is a measure preserving bijection from $\mathcal{S}_q$ to $\mathcal{S}_q^c$. Consider the space $[0,1) \times \mathcal{S}$ with probability measure $\lambda \times \nu$. Extend ξ_q to $[0,1) \times \mathcal{S}$ by defining $\xi_q(x, \phi) = (x, \xi_q(\phi))$.

Suppose there exists a consistent procedure for estimating each transformation T_ϕ in the strong topology. For each $\phi \in \mathcal{S}$, positive integer n and λ almost every $x \in [0,1)$, let $\hat{T}_\phi[x,n]$ be the estimate of T_ϕ produced from the sequence $\{x, T_\phi x, \ldots, T_\phi^{n-1}\}$. Define the set

$$G_n = \left\{(x,\phi) : \lambda\{y : \hat{T}_\phi[x,n](y) = T_\phi y\} > \frac{3}{4}\right\}.$$

Since the procedure is assumed to be consistent, $\lim_{n\to\infty} \lambda \times \nu(G_n) = 1$. Choose N such that

$$\lambda \times \nu(G_N) > \frac{15}{16}.$$

Let q be such that $2^{q-1} > N$, and let B denote the union of the bottom 2^{q-1} levels of C_{q+2}. Then $\lambda(B) = \frac{1}{4}$, and hence

$$\lambda \times \nu(G_N \cap (B \times \mathcal{S}_q)) > \lambda(B)\nu(\mathcal{S}_q) - \frac{1}{16} = \frac{1}{16}.$$

Also,

$$(\lambda \times \nu)(G_N \cap (B \times \mathcal{S}_q^c)) > \frac{1}{16}.$$

Since the map ξ_q is measure preserving,

$$\lambda \times \nu(\xi_q(G_N \cap (B \times \mathcal{S}_q)) \cap G_N \cap (B \times \mathcal{S}_q^c)) > 0.$$

Therefore there exist $x \in B$ and $\phi \in \mathcal{S}_q$ such that both $(x,\phi) \in G_N$ and $(x, \xi_q(\phi)) \in G_N$. But this contradicts the consistency of the procedure since $T_\phi^i x = T_{\xi_q(\phi)}^i x$ for $i = 0, 1, \ldots, N-1$, but $d(T_\phi, T_{\xi_q(\phi)}) \geq 1/2$. $\square$

Remark: As shown by one of the referees, one may establish a similar counterexample based on a family $\mathcal{T}$ of transformations of the set $\Omega = [1] \times [2] \times [3] \times \cdots$, where $[n] = \{1, 2, \cdots, n\}$. The family $\mathcal{T}$ contains all infinite products of cyclic permutations of the coordinates of Ω.

Acknowledgements: The authors wish to thank the referees for their helpful comments and suggestions.

References

[1] Arnoux, Ornstein, Weiss, *Cutting and Stacking, Interval Exchanges and Geometric Model.* Israel Journal of Mathematics, 50 (1985), 160-168.

[2] T.A. Bick *On orbits in spaces of infinite measure.* Duke Mathematics Journal, 34(4) (1967), 717-724.

[3] T.A. Bick and J. Coffey *A class of examples of D-sequences.* Ergodic Theory & Dynamical Systems 11 (1991), 1-6.

[4] D. Bosq, and D. Guégan, *Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system.* Statistics and Probability Letters 25 (1995), 201–212.

[5] M. Casdagli, *Nonlinear prediction of chaotic time series.* Physica D, 35 (1989), 335-356.

[6] M. Casdagli, *Chaos and deterministic* versus *stochastic non-linear modeling.* Journal of the Royal Statistical Society, series B, 54 (1992), 303-328.

[7] J. Coffey *Uniformly distributed D-sequences.* In "Measure and measurable dynamics", Contemporary Mathematics 94 (1989), 97-112.

[8] J.D. Farmer and J.J. Sidorowich, *Predicting chaotic time series.* Physical Review Letters 59 (1987), 845-848.

[9] D.A. Kappos and S. Papadopoulou, *Transformations determined by the orbit of a general point.* Rev. Roum. Math. Pures et Appl. XII (9) (1967), 1297-1303.

[10] E.J. Kostelich and J.A. Yorke, *Noise reduction: finding the simplest dynamical system consistent with the data.* Physica D 41 (1990), 183-196.

[11] Lalley, S.P. (1999) Beneath the noise, chaos. *Annals of Statistics*, 27 461-479.

[12] Z.-Q. Lu and R.L. Smith, *Estimating local Lyapunov exponents.* Fields Institute Communications 11 (1997), 135–151.

[13] D. Maharam, On orbits under ergodic measure-preserving transformations. Transactions of the American Mathematical Society, 119 (1965), pp.51-66.

[14] A.B. Nobel, *First order predictive sequences and induced transformations.* Technical report, Department of Statistics, University of North Carolina, Chapel Hill, 1999.

[15] A.B. Nobel and T.M. Adams *On regression estimation from ergodic samples with additive noise* Submitted for publication, (2000).

[16] D. Nychka, S. Ellner, A.R. Gallant, and D. McCaffrey, *Finding chaos in noisy systems.* Journal of the Royal Statistical Society, series B, 54 (1992), 399-426.

[17] D.S. Ornstein and B. Weiss, *How Sampling Reveals a Process.* Annals of Probability, 18 (1990), 905-930.

[18] H.L. Royden, *Real Analysis*, third edition. Prentice Hall, New Jersey (1988).

[19] Y. Sun, *Isomorphisms for convergence structures.* Advances in Mathematics 116 (1995), 322-355.

[20] H. Tong, *Non-linear Time Series: a Dynamical System Approach.* Oxford University Press, (1990).