Ergod. Th. & Dynam. Sys. (1997), **17**, 941–956
Printed in the United Kingdom © *1997 Cambridge University Press*

Are the dimensions of a set and its image equal under typical smooth functions?*

TIMOTHY D. SAUER† and JAMES A. YORKE‡

† *Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA*

‡ *Institute of Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

(*Received* 15 *December* 1994 *and revised* 10 *August* 1996)

Abstract. We examine the question whether the dimension D of a set or probability measure is the same as the dimension of its image under a typical smooth function, if the range space is at least D-dimensional. If μ is a Borel probability measure of bounded support in $\mathbb{R}^n$ with correlation dimension D, and if $m \geq D$, then under almost every continuously differentiable function ('almost every' in the sense of prevalence) from $\mathbb{R}^n$ to $\mathbb{R}^m$, the correlation dimension of the image of μ is also D. If μ is the invariant measure of a dynamical system, the same is true for almost every delay coordinate map. That is, if $m \geq D$, then m time delays are sufficient to find the correlation dimension using a typical measurement function. Further, it is shown that finite impulse response (FIR) filters do not change the correlation dimension. Analogous theorems hold for Hausdorff, pointwise, and information dimensions. We show by example that the conclusion fails for box-counting dimension.

1. *Introduction*

Let S be a bounded Borel subset of $\mathbb{R}^n$, and let $G : \mathbb{R}^n \to \mathbb{R}^m$ be a 'typical' continuously differentiable function. Assume $\dim(S) \leq m$. In this paper we investigate whether

$$\dim G(S) = \dim(S),$$

or in other words, whether G preserves the dimension of S. The answer to such a question obviously depends on both the definition of dimension and the notion of typical.

Mattila [**10**], building on ideas of Kaufman [**8**] and Marstrand [**9**], showed that almost every orthogonal projection from $\mathbb{R}^n$ to $\mathbb{R}^m$ (in the sense of Haar measure on the orthogonal group) preserves the Hausdorff dimension of any set whose Hausdorff dimension is no greater than m. Our goals are to extend this result to general smooth

* Research supported in part by the National Science Foundation (Computational Mathematics and Physics) and the Department of Energy.

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, and to consider a wider range of definitions of dimension. For example, we will show that typical smooth functions preserve correlation dimension, and that Hausdorff dimension can be defined in terms of correlation dimension, recovering Mattila's result. One difficulty with extending such ideas is to determine a reasonable concept of 'typical' for general smooth functions, where there is no generalization of Haar measure. For this purpose we will often use the concept of prevalence [**7**].

We are primarily concerned with some standard metric dimensions such as Hausdorff, correlation, information, and box-counting dimension. For each of the first three types of dimensions, we show that $\dim(G(S)) = \dim(S)$ for a dense subset of functions G in the C^1 topology. In fact, we can say that 'almost every' function in $C^1(\mathbb{R}^n, \mathbb{R}^m)$ preserves dimension in the following sense: if we denote by M an $m \times n$ matrix, then for *each* C^1-function $G : \mathbb{R}^n \to \mathbb{R}^m$, the function $(G + M)(x)$ preserves the dimension of S for (Lebesgue) almost every choice of entries in M. On the other hand, box-counting dimension is not preserved for all sets S. We will exhibit subsets $S \subset \mathbb{R}^n$ such that $\text{boxdim}(G(S)) < \text{boxdim}(S)$ for *every* continuously differentiable function $G : \mathbb{R}^n \to \mathbb{R}^m$.

The concept of 'almost every' mentioned here fits into a more general context, using the terminology of [**7**]. It is stated in terms of all probability measures on $X = C^1(\mathbb{R}^n, \mathbb{R}^m)$. (A probability measure on X is a Borel measure μ satisfying $\mu(X) = 1$.) A Borel subset Y of the function space $X = C^1(\mathbb{R}^n, \mathbb{R}^m)$ is called *shy* if there exists a probability measure m on X for which the measure of every translate of Y is zero. That is, Y is shy if $m(Y + x) = 0$ for every x in X. A *prevalent* set is a set whose complement is shy. In particular, prevalent implies dense.

In the cases described above, the probability measure can be taken as follows. Let C be the set of linear transformations represented by $m \times n$ matrices whose entries are bounded by one in absolute value. The set C can be viewed as an mn-dimensional cube, on which we can put Lebesgue measure m normalized such that $m(C) = 1$. Extend m trivially to all of X, so that $m(X \setminus C) = 0$. Let X' be the subset of X consisting of functions that do not preserve the dimension of S. We prove that for every G, the translated set of functions $X' + G$ has m-measure zero. This shows that X' is shy. Therefore, our affirmative conclusion for the preservation of Hausdorff, correlation, and information dimensions holds for 'almost every' C^1-function from $\mathbb{R}^n$ to $\mathbb{R}^m$, in that the subset of such functions is a prevalent set.

In §2 we prove the dimension preservation result for correlation dimension D_2. Correlation dimension is defined for a Borel probability measure μ of bounded support. Assume that μ is a Borel probability measure on $\mathbb{R}^n$ satisfying $D_2(\mu) \leq m$. Let $G(\mu)$ denote the measure on $\mathbb{R}^m$ induced by μ, that is, for each subset S of $\mathbb{R}^m$, $G(\mu)(S) \equiv \mu(G^{-1}(S))$. We show that $D_2(G(\mu)) = D_2(\mu)$ for almost every smooth $G : \mathbb{R}^n \to \mathbb{R}^m$. Our proofs follow the general line of reasoning used by Mattila for Hausdorff dimension in [**10**].

Whether dimension is preserved is a necessary question from the point of view of interpretation of scientific measurement. Assume repeated observations are being made of a system, whose state space is $\mathbb{R}^n$. Often n is large, and the state is observed indirectly by making physical measurements. Assume that each observation results in m real numbers,

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

representing m simultaneous and independent measurements. By definition of state, the outcomes of these measurements depend uniquely on the state. The observations define a measurement function $G : \mathbb{R}^n \to \mathbb{R}^m$. Let the probability of the location of the state be described by a probability measure μ on $\mathbb{R}^n$. Can the correlation dimension of the state probability measure μ be determined from the observations alone? A probability measure $G(\mu)$ on $\mathbb{R}^m$ is induced by the observations. If the function G formed by the measurements satisfies $D_2(G(\mu)) = D_2(\mu)$, then the information about the correlation dimension of μ can be ascertained through measurement.

In §3 we take up the subject of other dimensions. First we point out the fact that the Hausdorff dimension D_H of a bounded Borel subset S of $\mathbb{R}^n$ is equal to the supremum of the correlation dimensions of probability measures defined on subsets of S. The fact that Hausdorff dimension is preserved under almost every smooth function follows directly from this characterization and the results of §2. Secondly, an alteration of the proof from §2 shows that for a given invariant measure μ, pointwise dimension is preserved for μ-almost every point, which implies that the information dimension is also preserved, i.e. $D_1(G(\mu)) = D_1(\mu)$ for almost every C^1-function $G : \mathbb{R}^n \to \mathbb{R}^m$. To finish the section, we present an example showing that no such result holds for box-counting dimension. We construct a compact subset Q of $\mathbb{R}^{10}$ with box-counting dimension $\text{boxdim}(Q) = 5$, that has the property that $\text{boxdim}(G(Q)) < 4$ for every continuously differentiable function $G : \mathbb{R}^{10} \to \mathbb{R}^6$. More generally, for any integers $0 < m < n$, and for any $d \in (0, n)$, there is a set $Q \subset \mathbb{R}^n$ satisfying $\text{boxdim}(Q) = d$ and $\text{boxdim}(G(Q)) \leq \overline{d} \equiv (d^{-1} + m^{-1} - n^{-1})^{-1} < d$ for every smooth function $G : \mathbb{R}^n \to \mathbb{R}^m$.

In §4 we point out that there is a considerable need to investigate dimension preservation for smooth functions $G : \mathbb{R}^n \to \mathbb{R}^m$ that are typical *within particular subspaces* of $C^1(\mathbb{R}^n, \mathbb{R}^m)$. For example, suppose $F : \mathbb{R}^n \to \mathbb{R}^n$ denotes a dynamical system with invariant measure μ. For each smooth $h : \mathbb{R}^n \to \mathbb{R}$, define the delay coordinate map $G_{F,h} : \mathbb{R}^n \to \mathbb{R}^m$ by

$$G_{F,h}(x) = (h(x), h(F^{-1}(x)), \ldots, h(F^{-(m-1)}(x))).$$

For a fixed dynamical system F, delay coordinate maps $G_{F,h}$ form a proper linear subspace of $C^1(\mathbb{R}^n, \mathbb{R}^m)$ that is often used by experimental scientists for reconstructing attractors from a single scalar time series of measurements [**6, 14, 15**]. We show in §4 that $D_2(G_{F,h}(\mu)) = D_2(\mu)$ for almost every smooth $h : \mathbb{R}^n \to \mathbb{R}$, with similar results for Hausdorff and information dimension.

Finally, we discuss results on the effect of finite impulse response (FIR) filters on dimension measurement. We show that if P is a fixed full rank matrix representing the FIR filter, then these dimensions are preserved by the filtered delay coordinate map $G_{P,F,h} := PG_{F,h}$ for almost every smooth h. Thus almost every FIR-filtered delay coordinate map preserves correlation, Hausdorff, and information dimension when the range space has dimension greater than or equal to the dimension of the set. This differs from the infinite impulse response (IIR) case, as has been noted experimentally and theoretically [**1**]. The significance of the results of this article for numerical calculation of correlation dimension is the subject of [**3, 13**].

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

2. *Correlation dimension*

2.1. *Definition of correlation dimension.* Let μ be a Borel probability measure on $\mathbb{R}^n$ with bounded support. The *correlation function* $C(\varepsilon)$ is defined to be the proportion of pairs of points (x, y) that lie within ε of one another in $\mathbb{R}^n$, weighted according to μ. That is, $C(\varepsilon) = (\mu \times \mu)(Q_\varepsilon)$, where

$$Q_\varepsilon = \{(x, y) : |x - y| < \varepsilon\}. \tag{2.1}$$

Definition 2.1. The *correlation dimension* [**6**] of the measure μ is

$$D = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}, \tag{2.2}$$

if the limit exists.

There is no guarantee that the limit (2.2) exists. Because of this, we will commonly work with the lower correlation dimension given by definition (2.3) which is an extension of (2.2) but that is defined for all measures in a natural way.

Definition 2.2. The *lower correlation dimension* of the measure μ is

$$D_2(\mu) = \sup \left\{ s : \int_0^\infty \frac{dC(\varepsilon)}{\varepsilon^s} < \infty \right\}. \tag{2.3}$$

The integral is defined because the correlation function $C(\varepsilon)$ is a nondecreasing positive function of ε, implying that $dC(\varepsilon)$ is defined except for a countable set of ε. Next we show that Definition 2.2 is an extension of the previous definition, by showing that it is found by replacing the limit in equation (2.2) with lim inf.

PROPOSITION 2.3.

$$D_2(\mu) = \liminf_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}. \tag{2.4}$$

In particular, if the limit in (2.2) *exists then the* $D_2(\mu)$ *of* (2.3) *also equals this limit.*

Proof. Let D denote the right-hand side of the equality. We will show $D = D_2$. The definition of lim inf implies that for any $\delta > 0$, there exists $\varepsilon_0 > 0$ such that

$$C(\varepsilon) < \varepsilon^{D-\delta} \quad \text{for } 0 < \varepsilon < \varepsilon_0, \tag{2.5}$$

and that there exists a sequence $\varepsilon_k \to 0$ such that $C(\varepsilon_k) > \varepsilon_k^{D+\delta}$. Given $0 \le s < D$, we show that the integral of (2.3) converges as follows. Set $\delta = (D - s)/2 > 0$, and let ε_0 be as above. Integration by parts yields

$$\int_0^\infty \frac{dC(\varepsilon)}{\varepsilon^s} = s \int_0^\infty \varepsilon^{-s-1} C(\varepsilon)\, d\varepsilon.$$

Since $\int_{\varepsilon_0}^\infty dC(\varepsilon)/\varepsilon^s < \infty$, we need only consider the remainder:

$$\int_0^{\varepsilon_0} \varepsilon^{-s-1} C(\varepsilon)\, d\varepsilon < \int_0^{\varepsilon_0} \varepsilon^{-s-1} \varepsilon^{D-\delta}\, d\varepsilon = \int_0^{\varepsilon_0} \varepsilon^{\delta-1}\, d\varepsilon = \frac{\varepsilon_0^\delta}{\delta} < \infty.$$

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

On the other hand, given $0 < D < s$, we show that the integral of (2.3) diverges. In this case, set $\delta = (s - D)/2 > 0$. Then for any ε_k from the sequence above,

$$\begin{aligned} \int_0^{\varepsilon_k} \frac{dC(\varepsilon)}{\varepsilon^s} &\geq \varepsilon_k^{-s} \int_0^{\varepsilon_k} dC(\varepsilon) \\ &= \varepsilon_k^{-s} C(\varepsilon_k) \\ &> \varepsilon_k^{-s} \varepsilon_k^{D+\delta} = \varepsilon_k^{-\delta}. \end{aligned} \tag{2.6}$$

Since ε_k can be chosen arbitrarily small, there is no upper bound for $\int_0^\infty \varepsilon^{-s}\, dC(\varepsilon)$. Thus $D = D_2$. □

The integral of (2.3) can be rewritten as

$$\langle |x - y|^{-s} \rangle_\mu \equiv \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|x - y|^s}\, d\mu \times d\mu, \tag{2.7}$$

where the notation $\langle \cdot \rangle_\mu$ denotes the average taken with respect to μ. In the usual terminology, the quantity $|x - y|^{-s}$ is called the *Riesz kernel*, and the average in (2.7) is called the *s-energy* of the measure μ. The correlation dimension of a measure is the supremum of s for which the measure has finite s-energy.

2.2. *Prevalence.* When referring to properties held by elements of infinite-dimensional metric linear spaces X, such as $X = C^1(\mathbb{R}^n, \mathbb{R}^m)$, we will use the term 'almost every' to mean that the set of elements with a given property is prevalent [**7**]. This use of the term 'almost every' was developed as an extension of Lebesgue almost every to the case of infinite-dimensional spaces. A Borel probability measure m on the complete linear space X is said to be *transverse* to a Borel subset $S \subset X$ if $m(S + x) = 0$ for every $x \in X$. A subset Y of X is *prevalent* if there exists a probability measure transverse to the complement of Y.

Prevalence is an alternative notion of genericity that seems appropriate when probabilistic properties are involved, which is common in dynamical systems theory. A property on a complete metric space is said to be *generic* if the set on which it holds is *residual*, meaning that it contains a countable intersection of open dense sets. Both prevalent sets in a complete metric space and residual sets (according to Baire) are dense. The two notions agree for some sets; for example, the set of points in $X = R^n$ with irrational coordinates is prevalent and residual. However, while a subset of R^n is prevalent if and only if it has full Lebesgue measure, a residual set can have Lebesgue measure zero (see [**7**] for some examples).

In this paper, X will represent either $C^1(\mathbb{R}^n, \mathbb{R}^m)$ or $C^1(\mathbb{R}^n, \mathbb{R})$, and Lebesgue measure on a finite-dimensional cube will be used for the measure transverse to the complement. For example, we will show that there is a finite-dimensional cube $C = \{\sum_{i=1}^k v_i G_i : |v_i| \leq 1\}$ of smooth functions from $\mathbb{R}^n \to \mathbb{R}^m$ such that for each smooth function $G_0 : \mathbb{R}^n \to \mathbb{R}^m$, the perturbed function $G = G_0 + \sum_{i=1}^k v_i G_i$ satisfies $D_2(G(A)) = D_2(A)$ for Lebesgue almost every $v = (v_1, \ldots, v_k)$ in the unit cube. That is, the subset of continuously differentiable functions which do not preserve dimension D_2 intersects the set $\{G_0 + \sum v_i G_i\}$ in a set of k-dimensional Lebesgue measure zero. Since this is true for every G_0, the set of non-dimension-preserving functions is transverse

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

to Lebesgue measure on the k-dimensional cube C. Therefore, the subset of $C^1(\mathbb{R}^n, \mathbb{R}^m)$ which preserves correlation dimension is prevalent. In particular, prevalent implies dense, so the theorem also holds in the weaker sense of a dense subset of smooth functions G.

2.3. *Preservation of correlation dimension.*

Notation 2.4. In $\mathbb{R}^m$, we denote by $B_m(r)$ the m-dimensional ball of radius r centered at 0, and by $C_m(r)$ the m-dimensional cube $\{x : |x_i| \le r, i = 1, \ldots, m\}$ centered at 0.

To prove the main theorems, we need the following well-known fact about integrating across singularities, and a couple of other propositions.

LEMMA 2.5. *The integral*

$$I_{s,m} = \int_{B_m(1)} \frac{dy}{|y|^s}$$

converges if and only if $s < m$.

LEMMA 2.6. *Let A be an $m \times k$ matrix, $0 \le r$, $0 \le s < m \le k$, and let $b \in \mathbb{R}^k$. If $AB_k(1) \supset B_m(\delta)$, then*

$$\int_{C_k(r)} \frac{dx}{|Ax + b|^s} \le \frac{2^{k-m}(r\sqrt{k})^{k-m-s}}{\delta^s} \int_{B_m(1)} \frac{dy}{|y|^s}. \tag{2.8}$$

Proof. We may assume $b = 0$, since

$$\int_{C_k(r)} \frac{dx}{|Ax + b|^s} \le \int_{C_k(r)} \frac{dx}{|Ax|^s}$$

for $b \in \mathbb{R}^k$. The singular value decomposition is $A = U\Sigma V^T$, where U and V are orthogonal matrices of shape $m \times m$ and $k \times k$, respectively, and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_m, 0, \ldots, 0)$, where $\sigma_1 \ge \cdots \ge \sigma_m$. The hypotheses imply that $\sigma_m \ge \delta$. Since A maps the m-dimensional ball $B_m(r)$ spanned by $\{v_1, \ldots, v_m\}$, the first m columns of V, to an ellipsoid with axes $r\sigma_i$, we have

$$\int_{B_m(r)} \frac{dx}{|Ax|^s} \le \frac{1}{r^s \sigma_m^s} \int_{B_m(1)} \frac{dy}{|y|^s}.$$

Let $C_{k-m}(r)$ be the $(k-m)$-dimensional cube of side length $2r$, centered at 0, spanned by the remaining columns of V. It follows that

$$\int_{B_k(r)} \frac{dx}{|Ax|^s} \le \int_{B_m(r)\times C_{k-m}(r)} \frac{dx}{|Ax|^s} \le \frac{(2r)^{k-m}}{r^s \sigma_m^s} \int_{B_m(1)} \frac{dy}{|y|^s}.$$

The fact that

$$\int_{C_k(r)} \frac{dx}{|Ax|^s} \le \int_{B_k(r\sqrt{k})} \frac{dx}{|Ax|^s}$$

completes the proof. □

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

If μ is a measure on $\mathbb{R}^n$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ is a function, define the *induced measure* $G(\mu)$ on $\mathbb{R}^m$ by

$$G(\mu)(S) = \mu(G^{-1}(S)) \tag{2.9}$$

for each subset S of $\mathbb{R}^m$. In this section we are interested in showing that for the correlation dimension D_2 defined above, the equality $D_2(G(\mu)) = D_2(\mu)$ holds for almost every G, in an appropriate sense.

The next result is the main proposition of the section. It gives a criterion for establishing dimension preservation for smooth functions $G : \mathbb{R}^n \to \mathbb{R}^m$ that are typical within particular linear subspaces $L(V)$ of $C^1(\mathbb{R}^n, \mathbb{R}^m)$. The first case of interest is $V = C^1(\mathbb{R}^n, \mathbb{R}^m)$, where L is the identity function. A second case, pursued in §4, assumes an underlying smooth invertible dynamical system $F : \mathbb{R}^n \to \mathbb{R}^n$. For each smooth $h : \mathbb{R}^n \to \mathbb{R}$, define the *delay coordinate map* $G_{F,h} : \mathbb{R}^n \to \mathbb{R}^m$ by

$$G_{F,h}(x) = (h(x), h(F^{-1}(x)), \ldots, h(F^{-(m-1)}(x))).$$

Then setting $V = C^1(\mathbb{R}^n, \mathbb{R})$, the identification $L(h) = G_{F,h}$ defines a proper linear subspace $L(V)$ of $C^1(\mathbb{R}^n, \mathbb{R}^m)$, the subspace of delay coordinate maps. Our conclusion in the second case will be that for a fixed F, the delay coordinate map $G_{F,h}$ preserves dimension for almost every h.

PROPOSITION 2.7. *Let μ be a probability measure with bounded support on $\mathbb{R}^n$ which satisfies $D_2(\mu) \leq m$. Let V be a complete metric linear space, and let $L : V \to C^1(\mathbb{R}^n, \mathbb{R}^m)$ be linear. Assume that there exists a k-dimensional unit cube $C = C_k(1) \subset V$ and a constant $K > 0$ such that*

$$\{Lv(x) - Lv(y) \mid v \in C\} \supset B_m(K|x - y|) \tag{2.10}$$

for each pair x, y in $\mathbb{R}^n$. Then for any $G_0 \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, the function $G_v = G_0 + Lv$ satisfies $D_2(G_v(\mu)) = D_2(\mu)$ for (Lebesgue) almost every $v \in C$.

Proof. First note that $D_2(G(\mu)) \leq D_2(\mu)$ for any Lipschitz function $G : \mathbb{R}^n \to \mathbb{R}^m$. In fact, for each pair x, y in the support of μ, the inequality $|G(x) - G(y)| \leq \lambda|x - y|$ holds. We will consider expression (2.7) for the integral in (2.3). The average value of $|G(x) - G(y)|^{-s}$ is bounded below by the average value of $|x - y|^{-s}/\lambda^s$, where the average is taken with respect to the measure $\mu \times \mu$. If $s > D_2(\mu)$, then both averages diverge, so $s \geq D_2(G(\mu))$.

For each $G_0 \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ and $v \in C$, define $G_v = G_0 + Lv$. It remains to show that if $D_2(\mu) \leq m$, then for each G_0 and for almost every $v \in C$, the reverse inequality $D_2(\mu) \leq D_2(G_v(\mu))$ holds. First, we note that the hypotheses together with Lemma 2.6 imply the inequality

$$\int_C \frac{dv}{|G_v(x) - G_v(y)|^s} \leq \frac{(2\sqrt{k})^{k-m} K^{-s}}{|x - y|^s} \int_{B_m(1)} \frac{dr}{|r|^s} \tag{2.11}$$

for each pair $x \neq y$ in $\mathbb{R}^n$ and every $0 \leq s < m$. In applying Lemma 2.6 we have set $A : C \to \mathbb{R}^m$ to be the linear transformation $v \mapsto L(v)(x) - L(v)(y)$, and

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

$b = G_0(x) - G_0(y)$. Now let $s < D_2(\mu)$. By reversing the order of integration, we see that the integral

$$\int_C \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{d\mu \times d\mu}{|G_v(x) - G_v(y)|^s}\, dv \tag{2.12}$$

is bounded above by

$$(2\sqrt{k})^{k-m} K^{-s} I_{s,m} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{d\mu \times d\mu}{|x - y|^s},$$

which is finite by the definition of $D_2(\mu)$. It follows that for almost every $v \in C$, the inner integral

$$\langle |G_v(x) - G_v(y)|^{-s} \rangle_\mu = \langle |x - y|^{-s} \rangle_{G_v(\mu)}$$

of (2.12) must be finite, so that $s \leq D_2(G_v(\mu))$. Since this is true for a sequence of s approaching $D_2(\mu)$, we conclude that $D_2(G_v(\mu)) = D_2(\mu)$ for almost every $v \in C$. □

THEOREM 2.8. *Let μ be a probability measure with bounded support on $\mathbb{R}^n$ which satisfies $D_2(\mu) \leq m$. Let $C_{mn}(1)$ denote the set of $m \times n$ matrices whose entries are bounded by one in absolute value. Then:*

1. *for any $G_0 \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, the function $G_M \equiv G_0 + M$ satisfies $D_2(G_M(\mu)) = D_2(\mu)$ for (Lebesgue) almost every $M \in C_{mn}(1)$, where $G_M(\mu)$ denotes the measure induced on $\mathbb{R}^m$ by G_M; and*
2. *for almost every C^1 function $G : \mathbb{R}^n \to \mathbb{R}^m$, the equality $D_2(G(\mu)) = D_2(\mu)$ holds.*

Proof. (2) follows from (1) by the definition of prevalence. To prove (1), we will apply Proposition 2.7 with $V = C^1(\mathbb{R}^n, \mathbb{R}^m)$ and L the identity mapping. Let M_{ij} denote the linear transformation represented by an $m \times n$ matrix with (i, j) entry 1 and all other entries zero. Define the unit cube $C_{mn}(1) = \left\{ \sum_{i=1}^m \sum_{j=1}^n v_{ij} M_{ij} : |v_{ij}| \leq 1 \right\}$.

For each pair x, y in $\mathbb{R}^n$, one of the entries of $x - y$ has magnitude at least $|x - y|/\sqrt{n}$. Since the functions M_{ij} are linear, the set $\{M(x - y) : M \in C_{mn}(1)\}$ covers $B_m(0, |x - y|/\sqrt{n})$. Thus Proposition 2.7, with $K = 1/\sqrt{n}$, shows that for each $G_0 \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, the function $G_M = G_0 + M$ satisfies $D_2(G_M(\mu)) = D_2(\mu)$ for (Lebesgue) almost every $M \in C_{mn}(1)$. □

3. *Other dimensions*

3.1. *Hausdorff dimension.* Next we show that Theorem 2.8 remains true if correlation dimension is replaced with Hausdorff dimension. The statement is amended to concern sets rather than measures.

Definition 3.1. For a Borel subset A of $\mathbb{R}^n$, an *ε-cover* of A is a collection of open balls U_i of diameter $|U_i|$ at most ε whose union contains A. The *Hausdorff dimension* of A is the number $D_H(A)$ satisfying

$$\liminf_{\varepsilon \to 0} \sum_{i=1}^{\infty} |U_i|^s = \begin{cases} \infty & \text{if } s < D_H(A) \\ 0 & \text{if } s > D_H(A), \end{cases}$$

where the inf is taken over all ε-covers.

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

Alternatively, the Hausdorff dimension $D_H(A)$ is the supremum of the real numbers s satisfying

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{|x-y|^s}\, d\mu \times d\mu < \infty, \tag{3.1}$$

where the supremum is taken not only over s but also over all probability measures μ supported on A [**5**, Theorem 4.13]. This means that the Hausdorff dimension of a Borel set can be defined in terms of lower correlation dimension:

$$D_H(A) = \sup_{\mu}\{D_2(\mu) : \mu(\mathbb{R}^n) = \mu(A)\}. \tag{3.2}$$

Therefore the proof that smooth functions preserve Hausdorff dimension follows from the correlation dimension case.

THEOREM 3.2. *Let A be a Borel subset of $\mathbb{R}^n$ which satisfies $D_H(A) \le m$. Let $C_{mn}(1)$ denote the set of $m \times n$ matrices whose entries are bounded by one in absolute value. Then:*

1. *for any $G_0 \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, the function $G_M \equiv G_0 + M$ satisfies $D_H(G_M(A)) = D_2(A)$ for (Lebesgue) almost every $M \in C_{mn}(1)$; and*
2. *for almost every C^1-function $G : \mathbb{R}^n \to \mathbb{R}^m$, the equality $D_H(G(A)) = D_H(A)$ holds.*

Proof. Hausdorff dimension cannot increase under a Lipschitz function. On the other hand, by the definition (3.2) of Hausdorff dimension in terms of correlation dimension, there is a countable sequence μ_i of probability measures such that $D_2(\mu_i) > D_H(A) - 1/i$. Theorem 2.8 says that for each i, the equality

$$D_2((G+M)(\mu_i)) = D_2(\mu_i) \tag{3.3}$$

holds for all linear transformations $M \in C_{mn}(1)$ not belonging to an exceptional set E_i of mn-dimensional Lebesgue measure zero. For M chosen outside of the measure zero subset consisting of the countable union of the E_i, we see that (3.3) holds for all i, which implies $D_H(G_M(A)) = D_2(A)$. The prevalence statement (2) follows from (1) by definition. □

3.2. *Information and pointwise dimension.* The statements of the previous section also hold for information dimension D_1, as a consequence of the fact that they hold for pointwise dimension. Assume that μ is a Borel probability measure on $\mathbb{R}^n$.

Definition 3.3. For each $x \in \mathbb{R}^n$, set $B(x, \varepsilon) = \{y : |x-y| < \varepsilon\}$, and define the *pointwise dimension* of μ at x by

$$d_x(\mu) = \lim_{\varepsilon \to 0} \frac{\log \mu(B(x,\varepsilon))}{\log \varepsilon}, \tag{3.4}$$

when the limit exists.

Whether the above limit exists or not, we will in addition define the number $d'_x(\mu)$ satisfying

$$\int_{\mathbb{R}^n} \frac{d\mu(y)}{|x-y|^s} = \begin{cases} \infty & \text{if } s > d'_x(\mu) \\ <\infty & \text{if } s < d'_x(\mu). \end{cases} \tag{3.5}$$

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

When the pointwise dimension $d_x(\mu)$ exists, then $d_x(\mu) = d'_x(\mu)$. The proof of this fact is similar to the proof of Proposition 2.3 and is omitted.

The information dimension [**12**] measures the rate of growth of the information given by successively finer discretizations of the state space.

Definition 3.4. Let X be a finite partition of the space, and define

$$H_\mu(X) = -\sum_{S\in X} \mu(S)\log\mu(S)$$

and

$$H_\mu(\varepsilon) = \inf_{\|X\|\le\varepsilon} H_\mu(X).$$

The *information dimension* of μ is

$$D_1(\mu) = \lim_{\varepsilon\to 0} \frac{H_\mu(\varepsilon)}{\log 1/\varepsilon}.$$

Young [**16**] showed that under the assumption that the pointwise dimension at x is the same for μ-almost every x, this common value equals the information dimension $D_1(\mu)$. Eckmann and Ruelle [**4**] conjecture that this assumption holds for a sufficiently smooth diffeomorphism on a compact Riemannian manifold with an ergodic Borel invariant measure and nonzero Lyapunov exponents. Young [**16**] proved this conjecture for two-dimensional manifolds.

THEOREM 3.5. *Let μ be a probability measure with bounded support on $\mathbb{R}^n$ for which the pointwise dimension $d_x(\mu)$ is a constant D on a set of full μ-measure, and such that $D \le m$. For almost every C^1-function $G : \mathbb{R}^n \to \mathbb{R}^m$, if the pointwise dimension exists on a set of full $G(\mu)$ measure, then $d_y(G(\mu)) = D$ for $G(\mu)$ almost every y, and $D_1(G(\mu)) = D_1(\mu) = D$.*

Proof. This proof follows along the lines of Proposition 2.7 and Theorem 2.8, except that double integrals for correlation dimension are replaced by single integrals for pointwise dimension. For μ almost every point x in the domain, we have $d_x = d'_x = D = D_1(\mu)$ (see equation 3.5), so the integral

$$\int_{\mathbb{R}^n} \frac{d\mu(y)}{|x-y|^s} \tag{3.6}$$

is finite for s less than D. Define the set A_L to be the set of x for which the integral (3.6) is less than the positive integer L. By integrating over $A_L \times \mathbb{R}^n$ (instead of $\mathbb{R}^n \times \mathbb{R}^n$ as in the proof of Proposition 2.7), we find that

$$\begin{aligned}\int_C \int_{A_L\times\mathbb{R}^n} \frac{d\mu(x)\times d\mu(y)}{|G_v(x)-G_v(y)|^s}\,dv &\le (2\sqrt{k})^{k-m} K^{-s} I_{s,m} \int_{A_L\times\mathbb{R}^n} \frac{d\mu(x)\times d\mu(y)}{|x-y|^s} \\ &\le (2\sqrt{k})^{k-m} K^{-s} I_{s,m} L, \end{aligned} \tag{3.7}$$

where $k = mn$, and where $C = C_{mn}(1)$, G_v, and K are as in the proof of Proposition 2.7. We conclude that for $s < D$ and for Lebesgue almost every $v \in C$, the integral

$$\int_{A_L} \int_{\mathbb{R}^n} \frac{d\mu(y)}{|G_v(x)-G_v(y)|^s}\,d\mu(x)$$

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

is finite for all L. Since $\mu(\cup_{L=1}^{\infty} A_L) = \mu(\mathbb{R}^n)$, we have for Lebesgue almost every $v \in C$ that

$$\int_{\mathbb{R}^n} \frac{d\mu(y)}{|G_v(x) - G_v(y)|^s} < \infty$$

for μ almost every x. For these G_v, it follows that $G_v(\mu)$ almost every point x of $G(\mathbb{R}^n)$ satisfies $d'_x = D_1(\mu)$. If the pointwise dimension exists for $G_v(\mu)$ almost every point, then $d' = d$ for these points and the information dimension of the image is also $D_1(\mu)$. □

3.3. *Box-counting dimension.* The above dimension preservation theorems do not hold if box-counting dimension D_0 is used. We show this with the following example.

Definition 3.6. Write $N(A, \varepsilon)$ for the number of balls in the smallest ε-cover of A. Then we define the *box-counting dimension* of A as

$$D_0(A) = \limsup_{\varepsilon \to 0} \frac{\log N(A, \varepsilon)}{\log 1/\varepsilon}. \tag{3.8}$$

At times we will use the shorthand $N(A, \varepsilon) \approx \varepsilon^{-D_0(A)}$ as an exact equivalent for equation (3.8).

Example 3.7. We exhibit a compact subset Q of $\mathbb{R}^{10}$ with box-counting dimension $D_0(Q) = 5$, so that for *any* C^1-function $G : \mathbb{R}^{10} \to \mathbb{R}^6$, $D_0(G(Q)) < 4$. The set Q is countable, but the aspect of Q that is crucial is not the countability but the fact that a cover of ε-balls results in a very dense packing in some regions. Cantor sets could be constructed with similar nonhomogeneity properties resulting in images in certain spaces necessarily having lower box-counting dimension than the original set.

Let $0 < d < n$ be integers, and define $\alpha = (n/d) - 1$. In $\mathbb{R}^n$, for $k > 0$, define Q_k to be a set of points on the sphere centered at 0 of radius $1/k^\alpha$ with the property that no two points of Q_k are closer than $\alpha k^{-\alpha-1}$, and that subject to this restriction there are as many points in Q_k as possible. That means that the number of points in Q_k is $\approx k^{n-1}$ for large positive k. Define the compact set $Q = \{0\} \cup Q_1 \cup Q_2 \cup \cdots$. The spheres of radius $1/(k+1)^\alpha$ and $1/(k-1)^\alpha$ adjacent to Q_k have distances approximately $\alpha k^{-\alpha-1}$ from Q_k. Hence the distance between neighboring shells is about the same as the distance between nearest neighbors on those shells.

The box-counting dimension calculation for Q goes as follows. Let k be a positive integer and let $\varepsilon = \alpha k^{-\alpha-1}$. Because the interpoint distances inside radius $1/k^\alpha$ are smaller than ε, we need to cover virtually the entire solid disk D_k of radius $1/k^\alpha$. Covering this 'inner region' of Q requires $N(D_k, \varepsilon) \approx (k^\alpha/\varepsilon)^n \approx k^n \approx \varepsilon^{-d}$ balls. The remainder of the set, the 'outer region', needs one ball per point, since the interpoint distances are greater than ε. Thus $N(Q_1 \cup \cdots \cup Q_{k-1}, \varepsilon) \approx 1 + 2^{n-1} + \cdots + (k-1)^{n-1} \approx k^n \approx \varepsilon^{-d}$. The total is $N(Q, \varepsilon) \approx \varepsilon^{-d} + \varepsilon^{-d}$, so the box-counting dimension of Q is $D_0(Q) = d$.

Now let G be any smooth function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Without loss of generality we can assume that $G(0) = 0$, and multiplying G by a positive constant (thereby making no change in the dimension of the image), we assume that $|G(x)| \le |x|$ for $x \in Q$.

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

The sets Q_j for $j \geq k$ all map into D_k^m, the ball in $\mathbb{R}^m$ of radius $1/k^\alpha$, and so for the inner region, $N(D_k^m, \varepsilon) \approx (k^{-\alpha}/\varepsilon)^m$. The number of points in the outer region of Q (beyond radius $1/k^\alpha$) is proportional to k^n, and while many of these would be mapped into D_k^m, the number $k^n = \varepsilon^{-n/\alpha}$ is an upper bound on the number of points needed for an ε cover of these points.

Hence $N(G(Q), \varepsilon)$ is bounded by a number proportional to $(k^{-\alpha}/\varepsilon)^m + k^n \approx \varepsilon^{-\overline{d}}$. Choosing the balls so that this number is as small as possible means choosing $\overline{d}$ so that the two numbers of balls are equal, so $k^{-m\alpha}/\varepsilon^m \approx k^n \approx \varepsilon^{-\overline{d}}$ implies $\overline{d}^{-1} = d^{-1} + m^{-1} - n^{-1}$. The box-counting dimension of the image is $D_0(G(Q)) \leq \overline{d} = (d^{-1} + m^{-1} - n^{-1})^{-1} < d$. For example, the choice of $n = 10, m = 6$, and $d = 5$ will result in a set of dimension $D_0(Q) = 5$ whose image dimension under any C^1-function $G : \mathbb{R}^n \to \mathbb{R}^m$ is $D_0(G(Q)) < (5^{-1} + 6^{-1} - 10^{-1})^{-1} = 3.75$. Alternatively, Q could be a dimension 1.5 subset of $\mathbb{R}^3$, all of whose C^1 images in $\mathbb{R}^2$ have dimension less than 1.2.

4. *Delay coordinates*

4.1. *Dimension preservation for delay coordinate maps.* There are similar dimension preservation results for the subspace of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ which are delay coordinate maps. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ denote a smooth invertible dynamical system with invariant measure μ. For example, F may be the 'time-T' map which represents T time units of the flow of a system of differential equations. The probability measure μ is said to be *invariant* under F if $\mu(F^{-1}(S)) = \mu(S)$ for each subset S of $\mathbb{R}^n$. For each smooth $h : \mathbb{R}^n \to \mathbb{R}$, define the delay coordinate map $G_{F,h} : \mathbb{R}^n \to \mathbb{R}^m$ by

$$G_{F,h}(x) = (h(x), h(F^{-1}(x)), \ldots, h(F^{-(m-1)}(x))). \tag{4.1}$$

We begin with a polynomial interpolation lemma. Given interpolation nodes y_i and data z_i to be fitted, the lemma bounds the size of coefficients of the interpolating polynomials in terms of the distance between the interpolating nodes.

LEMMA 4.1. *Let $\{P_1, \ldots, P_{n_*}\}$ be the set of monomials in n variables of degree at most $2m-1$, where $n_* = \binom{2m-1+n}{n}$. Assume that $y_1, \ldots, y_{2m}$ lie in $\mathbb{R}^n$ and satisfy*

$$|y_i - y_j| \geq \begin{cases} \sigma & \text{if } |i-j| = m \\ \varepsilon & \text{if } |i-j| \neq m, \end{cases}$$

for some $\sigma, \varepsilon > 0$, and let $z_1, \ldots, z_{2m} \in \mathbb{R}$. Then there exist $c_1, \ldots, c_{n_*} \in \mathbb{R}$ such that

$$\sum_{k=1}^{n_*} c_k P_k(y_i) = z_i$$

for $1 \leq i \leq 2m$, and

$$|c_i| \leq \frac{2m \cdot \max\{|z_i|\} \max\{1, |y_i|\}^{2m-1}}{\sigma \varepsilon^{2m-2}}. \tag{4.2}$$

Proof. The Lagrange interpolating polynomial for the variable $y \in \mathbb{R}^n$ is

$$Q(y) = \sum_{j=1}^{2m} z_j \prod_{1 \leq i \leq 2m, i \neq j} \frac{(y - y_i)^T (y_j - y_i)}{|y_j - y_i|^2}.$$

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

Check that $Q(y_i) = z_i$ for $1 \leq i \leq 2m$. Each subproduct has form $(y - y_i)^T u_{ij}/|y_j - y_i|$, where u_{ij} is a unit vector. Each summand contains the product of $2m-1$ such terms. Of the $2m-1$ denominators $|y_j - y_i|$, one is bounded below by σ and $2m-2$ are bounded below by ε. Expanding $Q(x)$ into its monomial representation results in the bound (4.2) on $|c_i|$. □

To prove a dimension-preservation theorem about delay coordinate maps $G_{F,h}$ as in (4.1), it is necessary to put an extra condition on periodic orbits of low period for the diffeomorphism F. For example, if the invariant set of F consists entirely of fixed points, they will all be mapped to the diagonal $x_1 = \cdots = x_m$ under any delay coordinate map, and dimension cannot be preserved if the invariant set has dimension larger than one. To avoid pathological cases, we will assume that the set of period $\leq m$ points has μ-measure zero.

THEOREM 4.2. *Let μ be an invariant measure with bounded support under the diffeomorphism $F : \mathbb{R}^n \to \mathbb{R}^n$, such that periodic points of period $\leq m$ comprise a set of zero μ-measure. If $D_2(\mu) \leq m$, then for almost every C^1-function $h : \mathbb{R}^n \to \mathbb{R}$,*

$$D_2(G_{F,h}(\mu)) = D_2(\mu).$$

Proof. The proof is similar to that of Theorem 2.8. We want to show that for almost every perturbation $h = h_0 + \sum_{i=1}^{k} c_i h_i$ of the observation function h_0, the delay coordinate map $G_{F,h}$ preserves dimension.

The set of periodic points of period less than m is a Borel set, so we work on the full-measure complement S_0 of this set, which has the same dimension D_2 as μ. In the absence of periodic points of period less than m, we can assume the points $x, F^{-1}(x), \ldots, F^{-(m-1)}(x), y, \ldots, F^{-(m-1)}(y)$ are distinct. Then the role of the linear functions M_{ij} in the proof of Theorem 2.8 can be filled by interpolating polynomials $h_i : \mathbb{R}^n \to \mathbb{R}$.

Consider the subset of S_0 defined by $S_\varepsilon = \{x : |F^i(x) - F^j(x)| > \varepsilon, 0 \leq i < j < m\}$. This omits periodic orbits and nearly (within ε) periodic orbits of period less than m. The union of the S_ε for a sequence of $\varepsilon \to 0$ (say $\varepsilon = 1/j$) is S_0. We will show that a prevalent subset of observation functions h preserve dimension for each $\varepsilon > 0$. Since a countable intersection of prevalent subsets is prevalent, we will have shown that dimension is preserved on all of S_0.

Since F and F^{-1} are Lipschitz on the (bounded) support of μ, there exists $\delta \geq 1$ such that $\delta|x - y| \geq |F^i(x) - F^i(y)| \geq \delta^{-1}|x - y|$ for $0 \leq i < m$. For each pair of distinct points $x, y \in S_\varepsilon$, and for $0 \leq i \neq j < m$, note that

$$\begin{aligned} |F^i(x) - F^j(y)| &= |F^i(x) - F^j(x) + F^j(x) - F^j(y)| \\ &\geq ||F^i(x) - F^j(x)| - |F^j(x) - F^j(y)|| \\ &\geq \varepsilon - \delta|x - y|. \end{aligned}$$

Therefore, if we restrict consideration to pairs x, y within $\varepsilon/2\delta$ of one another, we conclude that $|F^i(x) - F^j(y)| \geq \varepsilon/2$ for $0 \leq i \neq j < m$. On the other hand, if $i = j$, we have only the bound $|F^i(x) - F^i(y)| \geq \delta^{-1}|x - y|$.

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

We will apply Proposition 2.7. Define $W = 2mR^{2m-1}/\varepsilon^{2m-2}$, where R is large enough so that the support of the measure μ is contained in a ball of radius R. We will use the k-dimensional cube $\sum c_i P_i$, where the P_i are monomials in n variables of degree at most $2m-1$, and $|c_i| \le 1$, for the set of functions. For x, y in S_ε, we can apply Lemma 4.1 with $y_1 = x$, $y_2 = F^{-1}(x), \ldots, y_m = F^{-(m-1)}(x)$, $y_{m+1} = y$, $y_{m+2} = F^{-1}(y), \ldots, y_{2m} = F^{-(m-1)}(y)$ to conclude that all points in the cube $C_m(|x-y|/\delta)$ can be written in the form $G_{F,\Sigma c_i P_i}(x) - G_{F,\Sigma c_i P_i}(y)$ for $|c_i| \le W$. Define the linear map $L : C^1(\mathbb{R}^n, \mathbb{R}) \to C^1(\mathbb{R}^n, \mathbb{R}^m)$ by $L(h) = G_{F,h}$. It follows that

$$\left\{ L\left(\sum_{i=1}^{k} c_i P_i\right)(x) - L\left(\sum_{i=1}^{k} c_i P_i\right)(y) : |c_i| \le 1 \right\} \supset B_m(|x-y|/\delta W). \tag{4.3}$$

Proposition 2.7 says that for any G_{F,h_0}, the perturbed delay coordinate map satisfies

$$D_2(G_{F,h_0+\Sigma c_i P_i}(\mu|_{A_\varepsilon})) = D_2(\mu|_{A_\varepsilon})$$

except for a Lebesgue measure zero subset of $\{c_i\}$. The union of these measure zero subsets over $\varepsilon = 1/j$ for $j = 1, 2, \ldots$ is again a measure zero exceptional set of $\{c_i\}$ outside of which

$$D_2(G_{F,h_0+\Sigma c_i P_i}(\mu)) = D_2(\mu). \qquad \square$$

4.2. *Filtered delay coordinate maps and extensions.* The filtered version, Theorem 4.3, is similar. The FIR filter P is an $m \times p$ matrix of rank m. For a function $h : \mathbb{R}^n \to \mathbb{R}$, the *filtered delay coordinate map* $G_{P,F,h} : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$G_{P,F,h}(x) = P[h(x), h(F^{-1}(x)), \ldots, h(F^{-(p-1)}(x))].$$

The conclusion is that correlation dimension of a measure μ is preserved for almost every h as long as $D_2(\mu) \le m$. Only minor modifications are needed in the proof of Theorem 4.2. The sets S_0 must avoid periodic points of period less than p (instead of m), and the linear map needed to apply Proposition 2.7 is $L : C^1(\mathbb{R}^n, \mathbb{R}) \to C^1(\mathbb{R}^n, \mathbb{R}^m)$ defined by $L(h) = G_{P,F,h}$. Using interpolation polynomials of degree at most $2p-1$ results in

$$\{G_{F,\Sigma c_i P_i}(x) - G_{F,\Sigma c_i P_i}(y) : |c_i| \le 1\} \supset B_p(|x-y|/\delta W) \tag{4.4}$$

for each pair x, y. Applying the matrix P to the left-hand side of equation (4.4) gives the necessary inclusion

$$\{G_{P,F,\Sigma c_i P_i}(x) - G_{P,F,\Sigma c_i P_i}(y) : |c_i| \le 1\} \supset PB_p(|x-y|/\delta W). \tag{4.5}$$

The result is the following.

THEOREM 4.3. *Let μ be an invariant measure under the diffeomorphism $F : \mathbb{R}^n \to \mathbb{R}^n$ whose periodic points of period $\le p$ comprise a set of zero μ-measure. Let P denote an $m \times p$ matrix of rank m. If $D_2(\mu) \le m$, then for almost every C^1-function h, $D_2(G_{P,F,h}(\mu)) = D_2(\mu)$.*

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

Remark 4.4. Several extensions of these results follow from straightforward changes in the above proofs. Delay coordinate maps and filtered delay coordinate maps preserve Hausdorff dimension and information dimension as well. The analogue of Theorem 3.2 for delay coordinate maps follows immediately from Theorem 4.2. It says that the Hausdorff dimension of an F-invariant set S is preserved as long as $D_H(S) \leq m$. The analogue of Theorem 3.5 for information dimension under delay coordinate maps follows from the proof with straightforward modifications. Adding the changes of Theorem 4.3 gives a theorem that typical FIR-filtered delay coordinate maps preserve Hausdorff and information dimension. Finally, a combination of the techniques of §2–4 of this paper can be used to show that the mixture $(G_0, G_{P_1,F,h_1}, \ldots, G_{P_j,F,h_j}) : \mathbb{R}^n \to \mathbb{R}^m$ of generic maps and generic (filtered) delay coordinate maps preserves the dimension D of any invariant measure/set subject to the assumption $D \leq m$.

Remark 4.5. In the case of smooth manifolds, there is previous work on the preservation of dimension under FIR filters. If A is a smooth (differentiable) compact manifold of dimension n (necessarily an integer), then the filtered delay coordinate map $G_{P,F,h} : A \to \mathbb{R}^m$ is a diffeomorphism for prevalent/generic $h : \mathbb{R}^n \to \mathbb{R}$, as long as $m > 2n$. This is proved in [**14**] and independently in [**2**] by a modification of Takens' proof [**15**]. Since a diffeomorphism on a compact manifold is bi-Lipschitz, correlation and other metric dimensions are preserved. One advantage of Theorem 4.3 is that it applies to sets that are not smooth manifolds. If the set (respectively, measure) in question is known to be contained in (respectively, supported on) an n-dimensional smooth manifold, then Theorem 4.3 says that m in the filtered delay coordinate map can be taken to be the smallest integer greater than or equal to the dimension of the set, which is smaller than n in general.

Acknowledgement. We wish to thank Yuval Peres for helpful discussions.

REFERENCES

[1] R. Badii, G. Broggi, B. Derighetti, M. Ravain, S. Ciliberto, A. Politi and M. A. Rubio. Dimension increase in filtered signals. *Phys. Rev. Lett.* **60** (1988), 979–982.

[2] D. S. Broomhead, J. P. Huke and M. R. Muldoon. Linear filters and nonlinear systems. *J. Roy. Stat. Soc.* B **54** (1992), 373.

[3] M. Ding, C. Grebogi, E. Ott, T. Sauer and J. A. Yorke. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? *Physica* **D69** (1993), 404–424.

[4] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57** (1985), 617–656.

[5] K. Falconer. *Fractal Geometry*. Wiley, 1990.

[6] P. Grassberger and I. Procaccia. Characterization of strange attractors. *Phys. Rev. Lett.* **50** (1983), 346–349.

[7] B. Hunt, T. Sauer and J. Yorke. Prevalence: a translation-invariant almost every for infinite dimensional spaces. *Bull. Amer. Math. Soc.* **27** (1992), 217–238. Prevalence: an addendum. *Bull. Amer. Math. Soc.* **28** (1993), 306–307.

[8] R. Kaufman. On Hausdorff dimension of projections. *Mathematika* **15** (1968), 153–155.

[9] J. M. Marstrand. Some fundamental geometrical properties of plane sets of fractional dimensions. *Proc. London Math. Soc.* **4** (1954), 257–302.

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press

[10] P. Mattila. Hausdorff dimension, orthogonal projections and intersections with planes. *Ann. Acad. Sci. Fennicae* A **1** (1975), 227–244.
[11] Y. Pesin. On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimensions. *J. Stat. Phys.* **71** (1993), 529–547.
[12] A. Rényi. Dimension, entropy and information. *Trans. Second Prague Conf. Information Theory, Statistical Decision Functions, Random Processes* (1957), 545–556.
[13] T. Sauer and J. A. Yorke. How many delay coordinates do you need? *Int. J. Bif. Chaos* **3** (1993), 737–744.
[14] T. Sauer, J. A. Yorke and M. Casdagli. Embedology. *J. Stat. Phys.* **65** (1991), 579–616.
[15] F. Takens. *Detecting Strange Attractors in Turbulence (Lecture Notes in Mathematics 898)*. Springer, 1981.
[16] L.-S. Young. Dimension, entropy and Lyapunov exponents. *Ergod. Th. & Dynam. Sys.* **2** (1982), 109–124.

https://doi.org/10.1017/S0143385797086252 Published online by Cambridge University Press