Embedding theory in geometric measure theory

Lecture notes for a minicourse at Uppsala University

Adam Śpiewak
Institute of Mathematics, Polish Academy of Sciences

Abstract

These notes accompany a minicourse on embedding theory for sets and measures in Euclidean spaces and their applications to dynamical systems and fractal geometry. The goal is to embed a set X (or a measure μ) into a Euclidean space $\mathbb{R}^k$, and the question is what information about the size, or complexity, of X is needed to guarantee such an embedding. The answer is given in terms of *dimension*: different notions of dimension correspond to different kinds of embeddings and to different regularity of the embedding. Motivated by the applications, we are interested in *generic* embeddings, i.e. in conditions under which a *typical* map from a given family of maps into $\mathbb{R}^k$ is injective (or injective on a set of full μ-measure). We discuss three instances: linear maps and orthogonal projections (Parts 1 and 2), time-delay coordinate maps for dynamical systems (Part 3) and natural projections of iterated function systems (Part 4). In all three cases the main tool is the *transversality method*. A recurring phenomenon is the *factor two*: deterministic embedding theorems require $k > 2 \dim X$, while in the probabilistic setting (injectivity on a set of full measure) $k > \dim \mu$ usually suffices.

Contents

Part 1: Basics of embedding theory **2**
1.1 Introduction . . . 2
1.2 A bit of history - topological and smooth embeddings. . . . 3
1.3 Geometric measure theory: dimensions of sets and measures . . . 4
1.4 Mañé's embedding theorem . . . 5
1.5 Mañé's theorem for measures . . . 9
1.6 Connections with projection theorems and conditional measures . . . 11
1.7 From projections to linear and Lipschitz maps; prevalence . . . 13
1.8 Why bother? On applications . . . 15

Part 2: Regularity of the embedding **16**
2.1 Regularity of the embedding: definitions . . . 16
2.2 The regularity theorem . . . 16
2.3 Proof of the regularity theorem for measures . . . 18
2.4 Bi-Lipschitz embeddings and the Assouad dimension . . . 21

Part 3: Dynamical systems — reconstruction and prediction from observed data **21**
3.1 Time-delay embeddings in time series analysis . . . 21
3.2 Takens-type embedding theorems . . . 22
3.3 The probabilistic approach: injectivity and predictability . . . 23
3.4 The Schroer–Sauer–Ott–Yorke conjectures . . . 29
3.5 Predictability and the Farmer–Sidorowich algorithm . . . 31

Version of October 27, 2026. Contact: aspiewak@impan.pl.

Part 4: Applications to fractal geometry **33**
4.1 Iterated function systems 33
4.2 Parametrized families of IFS 36
4.3 Multifractal analysis 38

Notation. We denote by $|\cdot|$ the Euclidean norm and by $\langle\cdot,\cdot\rangle$ the standard inner product on $\mathbb{R}^N$. $B(x,r)$ is the open ball of radius r centred at x (in $\mathbb{R}^N$, or in a metric space), and $B_N(0,1)$ is the unit ball of $\mathbb{R}^N$. Leb_N denotes the Lebesgue measure on $\mathbb{R}^N$.

Acknowledgement of AI use. These notes were prepared with the assistance of a large language model, used for LaTeX typesetting, language editing, literature search, compiling parts of the text based on sources selected by the author, and preparing the figures. The choice of material, the structure and the mathematical content were decided by the author, who takes full responsibility for the content of the notes.

Part 1: Basics of embedding theory

1.1 Introduction

The embedding problem. Our goal is to embed a set X (or a measure μ) into a Euclidean space $\mathbb{R}^k$. The basic question is:

> *What information about the size/complexity of X is needed to guarantee existence (or abundance) of embeddings of X into $\mathbb{R}^k$?*

The precise meaning of *embedding* is left vague for now and depends on the precise context. We will mean by it at least an injective map $\phi\colon X \to \mathbb{R}^k$, and in the best case scenario an isomorphism onto its image in the considered category (a homeomorphism for continuous maps, a bi-Lipschitz map for Lipschitz maps, a diffeomorphism for smooth maps etc.). For a measure μ, injectivity is required only on a set of full μ-measure. The answer will usually be given in terms of the *dimension* of X (or μ), and different notions of dimension will correspond to embeddings with certain properties.

Given a class $\mathcal{C}$ of maps $X \to \mathbb{R}^k$ (continuous, smooth, Lipschitz, linear, delay coordinate maps for observed dynamics, natural projections in fractal geometry, ...), there are two questions one can ask:

- *Existence:* is there an embedding $\phi\colon X \to \mathbb{R}^k$ in the class $\mathcal{C}$?
- *Genericity:* is a *typical* map $\phi \in \mathcal{C}$ an embedding?

Depending on the setting, "typical" can mean: in a dense G_δ subset of $\mathcal{C}$, for Lebesgue-almost every parameter of a finite-dimensional family, or for a prevalent set of maps (Definition 1.34).

In this course we will deal with the question of genericity, in the setting of geometric measure theory.

This is dictated by the applications we have in mind. In geometric measure theory we ask about almost every projection, with applications to analog compression by random linear maps (Parts 1 and 2); in time series analysis the observable is imposed by the experiment and we can at best hope that it is typical in some sense (Part 3); in fractal geometry we ask about typical parameters for families of iterated function systems (Part 4). The guiding questions are therefore:

(Q1) How large must k be, in terms of the dimension of X (or μ), for a typical map $\phi \in \mathcal{C}$ to be injective (resp. μ-a.s. injective)?

(Q2) How regular is the inverse ϕ^{-1}?

(Q3) Which notion of dimension governs the answers to (Q1)–(Q2)?

Dictionary between the parts. The following table summarizes how the same scheme appears in different parts of the course.

	Parts 1–2	Part 3	Part 4
parameter	$V \in \mathrm{Gr}(k, N)$	$\alpha \in \mathbb{R}^m$ (polynomial probes)	$\lambda \in U$
maps	orthogonal projection P_V	delay-coordinate map $\phi_{h_\alpha,k}$	natural projection π_λ
transversality	Lemma 1.15	Lemmas 3.9, 3.11	Definition 4.9
a.s. injectivity	$\dim_{\mathrm{H}} \mu < k$	$\dim_{\mathrm{H}} \mu < k$ + periodic points	$h(\mu)/\chi_{\mathcal{F}}(\mu) < d$
nearly bi-Lipschitz	$\dim_{\mathrm{A}} \operatorname{supp} \mu < k$	(smooth case: [Śpi25])	$s(\mathcal{F}) < d$

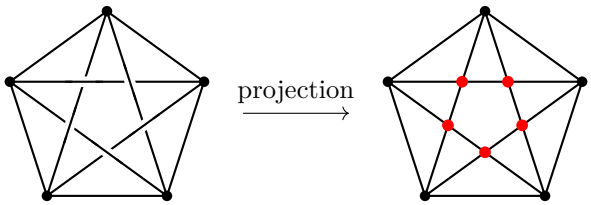
1.2 A bit of history - topological and smooth embeddings.

The first embedding theorem was proved in the topological setting. It allows to realize abstract metric spaces as subsets of $\mathbb{R}^k$.

Theorem 1.1 (Menger 1926, Nöbeling 1931)**.** *Let X be a compact metric space of Lebesgue covering (topological) dimension $\dim_{\mathrm{T}} X \le n$. Then the set of embeddings $X \to \mathbb{R}^{2n+1}$ (i.e. continuous maps $f : X \to \mathbb{R}^{2n+1}$ which are homeomorphisms onto their image) is a dense G_δ subset of $C(X, \mathbb{R}^{2n+1})$ with the supremum metric.*

Reference: [HW41, Thm. V.2], [Rob11, Thm. 1.12].

Example 1.2. The n-skeleton of the $(2n+2)$-dimensional simplex is an n-dimensional polyhedron which does not embed topologically into $\mathbb{R}^{2n}$ (for $n = 1$: the complete graph K_5 is not planar, see Figure 1). Hence $2n + 1$ in Theorem 1.1 is optimal. See [HW41, Example V.3], [Rob11, Section 1.3].



$K_5 \subset \mathbb{R}^3$ image in $\mathbb{R}^2$

Figure 1: The complete graph $X = K_5$, a compact metric space of dimension 1. Left: an embedding of K_5 into $\mathbb{R}^3$, as guaranteed by Theorem 1.1 (drawn in projection: at each crossing of the picture one edge passes above the other, so there are no self-intersections). Right: its projection to $\mathbb{R}^2$, with the self-intersections marked by red dots; by Example 1.2, no continuous map $K_5 \to \mathbb{R}^2$ is injective. There are only finitely many self-intersections, hence they form a set of measure zero with respect to the Lebesgue (length) measure on K_5. Maps which are injective outside a set of measure zero will be an important theme for much of the course.

Theorem 1.3 (Weak Whitney embedding theorem, 1936)**.** *Let M be a compact C^r manifold of dimension n, $r \ge 1$. Then C^r embeddings (i.e. C^r maps $f : M \to \mathbb{R}^{2n+1}$ which are C^r diffeomorphisms onto their images) form an open and dense subset of $C^r(M, \mathbb{R}^{2n+1})$.*

Reference: [Whi36]; see e.g. [Hir76, Ch. 2].

Theorem 1.4 (Strong Whitney embedding theorem, 1944)**.** *Let M be a C^r manifold of dimension $n \geq 1$, $r \geq 1$. Then M admits a C^r embedding into $\mathbb{R}^{2n}$.*

Reference: [Whi44], [Hir76, Ch. 2].

Remark 1.5. The importance of Whitney's theorems is that every abstract manifold can be realized as a submanifold of a Euclidean space: the intrinsic and the extrinsic definitions of a manifold describe the same objects. Theorem 1.3 is a genericity result, while Theorem 1.4 is an existence result only: a generic smooth map $M \to \mathbb{R}^{2n}$ (M compact) is an immersion with finitely many transverse double points, and such double points persist under small perturbations, so embeddings are not dense in $C^r(M, \mathbb{R}^{2n})$ [GG73, Ch. III, Section 3]. The dimension $2n$ is optimal in general: $\mathbb{RP}^2$ does not embed into $\mathbb{R}^3$, and more generally $\mathbb{RP}^n$ does not embed into $\mathbb{R}^{2n-1}$ if n is a power of 2 [MS74, Section 11]. For n not being a power of 2, every n-manifold embeds into $\mathbb{R}^{2n-1}$, and the minimal embedding dimension is not determined in general; it might be lower than $2n-1$ - see [Sko08, Section 2].

1.3 Geometric measure theory: dimensions of sets and measures

Our setting will be the one of geometric measure theory. As notions of dimension are crucial for the embedding theory, let us begin with introducing ones that we will use. Let X be a metric space.

Definition 1.6 (Hausdorff dimension)**.** For $s \geq 0$, the s-dimensional Haousdorff measure of X is

$$\mathcal{H}^s(X) = \lim_{\delta \to 0} \inf \Big\{ \sum_i \operatorname{diam}(U_i)^s : X \subset \bigcup_i U_i,\ \operatorname{diam}(U_i) \leq \delta \Big\},$$

and the Hausdorff dimension of X is $\dim_{\mathrm{H}} X = \inf\{s : \mathcal{H}^s(X) = 0\}$.

For $r > 0$, let $N(X, r) \in \mathbb{N} \cup \{\infty\}$ denote the minimal number of balls of radius r needed to cover X.

Definition 1.7 (Box-counting dimensions)**.** The lower and upper box-counting dimensions of X are

$$\underline{\dim}_{\mathrm{B}} X = \liminf_{r \to 0} \frac{\log N(X,r)}{-\log r}, \qquad \overline{\dim}_{\mathrm{B}} X = \limsup_{r \to 0} \frac{\log N(X,r)}{-\log r}.$$

Definition 1.8 (Assouad dimension)**.** The Assouad dimension of X is

$$\dim_{\mathrm{A}} X = \inf \Big\{ s \geq 0 :\ \exists\, M > 0\ \ \forall\, x \in X,\ 0 < r < R \quad N(X \cap B(x,R), r) \leq M \Big(\frac{R}{r}\Big)^s \Big\}.$$

Proposition 1.9 (Basic properties)**.** *(i) The following inequalities hold:*

$$\dim_{\mathrm{T}} X \leq \dim_{\mathrm{H}} X \leq \underline{\dim}_{\mathrm{B}} X \leq \overline{\dim}_{\mathrm{B}} X \leq \dim_{\mathrm{A}} X;$$

all these inequalities can be strict.

(ii) Stability:

- $\dim_{\mathrm{H}}$ *is countably stable:* $\dim_{\mathrm{H}} \big(\bigcup_{i=1}^{\infty} X_i\big) = \sup_i \dim_{\mathrm{H}} X_i$ *for any sets* $X_1, X_2, \ldots$,
- $\overline{\dim}_{\mathrm{B}}$ *is finitely stable:* $\overline{\dim}_{\mathrm{B}}(X \cup Y) = \max\{\overline{\dim}_{\mathrm{B}} X, \overline{\dim}_{\mathrm{B}} Y\}$,
- $\overline{\dim}_{\mathrm{B}} X = \overline{\dim}_{\mathrm{B}} \overline{X}$*; in particular* $\overline{\dim}_{\mathrm{B}}$ *is not countably stable (e.g.* $\overline{\dim}_{\mathrm{B}}(\mathbb{Q} \cap [0,1]) = 1$*),*
- $\dim_{\mathrm{A}}$ *is finitely stable and* $\dim_{\mathrm{A}} X = \dim_{\mathrm{A}} \overline{X}$.

(iii) If f is α-Hölder, then $\dim_{\mathrm{H}} f(X) \leq \dim_{\mathrm{H}} X/\alpha$ *and* $\overline{\dim}_{\mathrm{B}} f(X) \leq \overline{\dim}_{\mathrm{B}} X/\alpha$. $\dim_{\mathrm{A}}$ *is a bi-Lipschitz invariant, but it may* increase *under Lipschitz maps.*

(iv) For products:

- $\dim_{\mathrm{H}} X + \dim_{\mathrm{H}} Y \le \dim_{\mathrm{H}}(X \times Y)$,
- $\overline{\dim}_{\mathrm{B}}(X \times Y) \le \overline{\dim}_{\mathrm{B}} X + \overline{\dim}_{\mathrm{B}} Y$,
- $\dim_{\mathrm{A}}(X \times Y) \le \dim_{\mathrm{A}} X + \dim_{\mathrm{A}} Y$.

Reference: [Rob11, Ch. 2, 3, 9], [Fal14], [Mat95], [Fra20].

Definition 1.10 (Dimensions of measures)**.** Let μ be a finite Borel measure on $\mathbb{R}^N$. The lower and upper local dimensions of μ at x are

$$\underline{d}(\mu, x) = \liminf_{r \to 0} \frac{\log \mu(B(x,r))}{\log r}, \qquad \overline{d}(\mu, x) = \limsup_{r \to 0} \frac{\log \mu(B(x,r))}{\log r};$$

μ is *exact dimensional* if $\underline{d}(\mu, x) = \overline{d}(\mu, x) = \text{const}$ for μ-a.e. x. The (upper) and lower Hausdorff dimensions of μ are

$$\dim_{\mathrm{H}} \mu = \inf\{\dim_{\mathrm{H}} A : A \text{ Borel},\ \mu(\mathbb{R}^N \setminus A) = 0\},$$
$$\underline{\dim}_{\mathrm{H}} \mu = \inf\{\dim_{\mathrm{H}} A : A \text{ Borel},\ \mu(A) > 0\}.$$

Proposition 1.11.

$$\dim_{\mathrm{H}} \mu = \operatorname*{ess\,sup}_{x \sim \mu} \underline{d}(\mu, x) \quad \textit{and} \quad \underline{\dim}_{\mathrm{H}} \mu = \operatorname*{ess\,inf}_{x \sim \mu} \underline{d}(\mu, x).$$

Reference: [Fal97, Prop. 10.2–10.3].

Example 1.12. The dimension of a measure can be much smaller than the dimension of its support.

(i) For $\mu = \sum_n 2^{-n} \delta_{q_n}$ with $\{q_n\}$ dense in $[0,1]^N$ we have $\dim_{\mathrm{H}} \mu = 0$ and $\operatorname{supp} \mu = [0,1]^N$, so $\dim_{\mathrm{H}}(\operatorname{supp} \mu) = \dim_{\mathrm{B}}(\operatorname{supp} \mu) = N$.

(ii) For $p \in (0,1)$, let μ_p be the distribution of $\sum_{i=1}^{\infty} \xi_i 2^{-i}$, where $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = 0) = 1 - p$. Then $\operatorname{supp} \mu_p = [0,1]$, while μ_p is exact dimensional with

$$\dim_{\mathrm{H}} \mu_p = \frac{-p \log p - (1-p) \log(1-p)}{\log 2},$$

which is smaller than 1 for $p \neq \frac{1}{2}$ (for $p = \frac{1}{2}$, μ_p is the Lebesgue measure on $[0,1]$).

1.4 Mañé's embedding theorem

Let $1 \le k < N$ and let $\mathrm{Gr}(k, N)$ be the Grassmannian of k-dimensional linear subspaces of $\mathbb{R}^N$. It is a compact manifold of dimension $k(N-k)$, on which the orthogonal group $O(N)$ acts transitively. $\mathrm{Gr}(k, N)$ carries a unique $O(N)$-invariant Borel probability measure $\gamma_{k,N}$ [Mat95, Ch. 3]. Equivalently, $\gamma_{k,N}$ is the distribution of gV_0, where $V_0 \in \mathrm{Gr}(k, N)$ is fixed and $g \in O(N)$ is chosen according to the Haar probability measure, or the distribution of the span of k independent random vector distributied uniformly in $B_N(0,1)$. For instance, $\mathrm{Gr}(1,2)$ is the set of lines $\mathbb{R}(\cos\theta, \sin\theta)$, $\theta \in [0, \pi)$, and $\gamma_{1,2}$ corresponds to the normalized Lebesgue measure $d\theta/\pi$. For $V \in \mathrm{Gr}(k, N)$ let P_V denote the orthogonal projection onto V; identifying V with $\mathbb{R}^k$, we regard P_V as a linear map $\mathbb{R}^N \to \mathbb{R}^k$.

We now study the following problem:

> **Given $X \subset \mathbb{R}^N$, when is P_V injective on X for $\gamma_{k,N}$-almost every $V \in \mathrm{Gr}(k, N)$?**

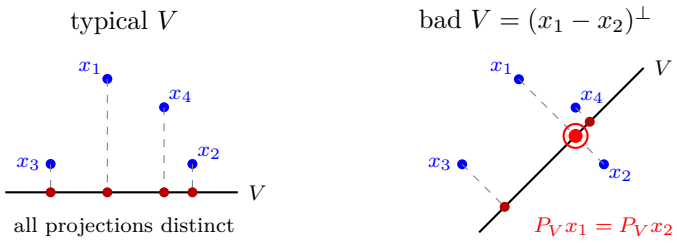


Figure 2: Orthogonal projections of finitely many points in $\mathbb{R}^2$ onto lines $V \in \mathrm{Gr}(1,2)$. Each pair $x_i \neq x_j$ has exactly one bad line $V = (x_i - x_j)^\perp$, on which $P_V x_i = P_V x_j$.

As a **warm-up**, we first consider the case of a countable set X.

Example 1.13 (Countable sets)**.** If $X \subset \mathbb{R}^N$ is countable, then P_V is injective on X for $\gamma_{k,N}$-a.e. $V \in \mathrm{Gr}(k,N)$ (for every $k \geq 1$).

Proof. Let

$$B = \{V \in \mathrm{Gr}(k,N) : P_V x = P_V y \text{ for some } x, y \in X,\ x \neq y\}$$

be the set of *bad* subspaces. We have to show that $\gamma_{k,N}(B) = 0$. Clearly

$$B = \bigcup_{\substack{x,y \in X \\ x \neq y}} B_{x,y}, \qquad \text{where} \quad B_{x,y} = \{V \in \mathrm{Gr}(k,N) : P_V x = P_V y\},$$

and the union is countable, since X is countable. It therefore suffices to show $\gamma_{k,N}(B_{x,y}) = 0$ for fixed $x, y \in X$ with $x \neq y$. Since $\ker P_V = V^\perp$, we have $P_V x = P_V y$ if and only if $x - y \in V^\perp$. Hence

$$B_{x,y} = \{V : x - y \in V^\perp\} = \{V : V \subset (x-y)^\perp\},$$

i.e. $B_{x,y}$ is the set of k-dimensional subspaces of the hyperplane $(x-y)^\perp$ (for $N = 2$, $k = 1$ this is a single line, $B_{x,y} = \{(x-y)^\perp\}$ - see Figure 2). In general, the set $B_{x,y}$ is a copy of $\mathrm{Gr}(k, N-1)$, a submanifold of $\mathrm{Gr}(k,N)$ of dimension $k(N-1-k) < k(N-k)$, hence $\gamma_{k,N}(B_{x,y}) = 0$. Therefore B is a countable union of $\gamma_{k,N}$-null sets, so $\gamma_{k,N}(B) = 0$. □

The proof shows that P_V is injective on X if and only if $V^\perp \cap (X - X) = \{0\}$. Hence controlling the size of the *difference set* $X - X = \{x - y : x, y \in X\}$ is crucial for the problem. Mañé's embedding theorem provides a solution.

Theorem 1.14 (Mañé's theorem)**.** *Let $X \subset \mathbb{R}^N$ satisfy $\mathcal{H}^k(X - X) = 0$; in particular it suffices that $\dim_{\mathrm{H}}(X - X) < k$ or $2\overline{\dim}_{\mathrm{B}} X < k$. Then P_V is injective on X for $\gamma_{k,N}$-a.e. $V \in \mathrm{Gr}(k,N)$.*

The proof of Theorem 1.14 is based on the following quantitative estimate.

Lemma 1.15 (Key estimate)**.** *For every $z \in \mathbb{R}^N \setminus \{0\}$ and $\varepsilon > 0$,*

$$\gamma_{k,N}(\{V \in \mathrm{Gr}(k,N) : |P_V z| \leq \varepsilon\}) \leq C_{N,k}\Big(\frac{\varepsilon}{|z|}\Big)^k,$$

where $C_{N,k} > 0$ depends only on N and k.

Proof. Since $|P_V z| \leq \varepsilon$ if and only if $|P_V(z/|z|)| \leq \varepsilon/|z|$, we may assume that $|z| = 1$. Since $\gamma_{k,N}$ is $O(N)$-invariant and $P_{gV}(gz) = gP_V z$ for $g \in O(N)$, we may moreover assume that $z = e_1 = (1, 0, \ldots, 0)$.

Step 1: a random basis. Let $l_1, \ldots, l_k$ be independent random vectors, each uniformly distributed in the unit ball $B_N(0,1)$, and let $V = \mathrm{span}(l_1, \ldots, l_k)$. Almost surely $l_1, \ldots, l_k$ are

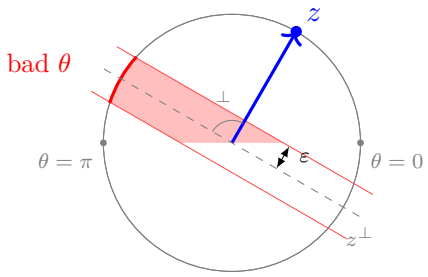


Figure 3: The key estimate for $k = 1$, $N = 2$ and $|z| = 1$. Write $V_\theta = \mathbb{R}e_\theta$, $e_\theta = (\cos\theta, \sin\theta)$, $\theta \in [0, \pi)$. Then $|P_{V_\theta} z| = |\langle e_\theta, z\rangle|$, so $|P_{V_\theta} z| \le \varepsilon$ if and only if e_θ lies in the strip of half-width ε around $z^\perp$. The set of such θ is an interval of length $2 \arcsin \varepsilon \le \pi\varepsilon$, hence $\gamma_{1,2}(\{V : |P_V z| \le \varepsilon\}) \le \varepsilon$.

linearly independent, so $V \in \mathrm{Gr}(k, N)$. For every $g \in O(N)$ the vectors $(gl_1, \dots, gl_k)$ have the same distribution as $(l_1, \dots, l_k)$, so the distribution of V is an $O(N)$-invariant Borel probability measure on $\mathrm{Gr}(k, N)$, hence it is equal to $\gamma_{k,N}$.

Step 2: reduction to k independent one-dimensional events. If $|P_V e_1| \le \varepsilon$, then, as $l_j \in V$ and $|l_j| \le 1$,

$$|\langle l_j, e_1\rangle| = |\langle l_j, P_V e_1\rangle| \le |P_V e_1| \le \varepsilon \qquad \text{for } j = 1, \dots, k.$$

Hence, by Step 1 and independence,

$$\gamma_{k,N}(\{V : |P_V e_1| \le \varepsilon\}) \le \mathbb{P}(|\langle l_j, e_1\rangle| \le \varepsilon \text{ for } j = 1, \dots, k) = \mathbb{P}(|\langle l_1, e_1\rangle| \le \varepsilon)^k. \tag{1}$$

Step 3: the one-dimensional estimate. The event $|\langle l_1, e_1\rangle| \le \varepsilon$ means that l_1 lies in the slab $\{l \in B_N(0,1) : |l^{(1)}| \le \varepsilon\} \subset [-\varepsilon, \varepsilon] \times B_{N-1}(0,1)$, where $l^{(1)}$ is the first coordinate of l. Therefore

$$\mathbb{P}(|\langle l_1, e_1\rangle| \le \varepsilon) \le \frac{2\varepsilon \,\mathrm{Leb}_{N-1}(B_{N-1}(0,1))}{\mathrm{Leb}_N(B_N(0,1))} = c_N\,\varepsilon.$$

Combining this with (1), we obtain $\gamma_{k,N}(\{V : |P_V e_1| \le \varepsilon\}) \le (c_N \varepsilon)^k$, so the lemma holds with $C_{N,k} = c_N^k$. □

Remark 1.16 (Transversality). Heuristically, as V varies, the point $P_V z$ moves in k independent directions with speed comparable to $|z|$, so the set of V with $|P_V z| \le \varepsilon$ has measure $\approx (\varepsilon/|z|)^k$ (see Figure 4 for $k = 1$, $N = 2$). Applied to $z = x - y$, this bounds the measure of the set of V for which $P_V x$ and $P_V y$ are ε-close. Estimates of this type are called *transversality conditions*. The name is explained by Figure 4: the graph of $\theta \mapsto P_\theta z$ crosses zero *transversally*, i.e. with nonzero slope $\pm|z|$, so it stays ε-close to zero only on a short interval of angles (applied to $z = x - y$: the graphs of $\theta \mapsto P_\theta x$ and $\theta \mapsto P_\theta y$ cross transversally). Transversality conditions will reappear for linear perturbations (Lemma 1.33), for delay maps (Lemma 3.9) and for natural projections of IFS (Definition 4.9).

Proof of Theorem 1.14. Assume first $\mathcal{H}^k(X - X) = 0$. As in the proof of Example 1.13, let

$$B = \{V \in \mathrm{Gr}(k, N) : P_V x = P_V y \text{ for some } x, y \in X,\ x \ne y\}$$

be the set of *bad* subspaces. We have to show that $\gamma_{k,N}(B) = 0$. Since $P_V x = P_V y$ if and only if $P_V(x - y) = 0$, and $(X - X) \setminus \{0\} = \bigcup_{n=1}^\infty Z_n$ with $Z_n = \{z \in X - X : |z| \ge 2^{-n}\}$, we have

$$B = \bigcup_{n=1}^\infty B_n, \qquad \text{where} \quad B_n = \{V \in \mathrm{Gr}(k, N) : P_V z = 0 \text{ for some } z \in Z_n\}.$$

It therefore suffices to show that $\gamma_{k,N}(B_n) = 0$ for every n.

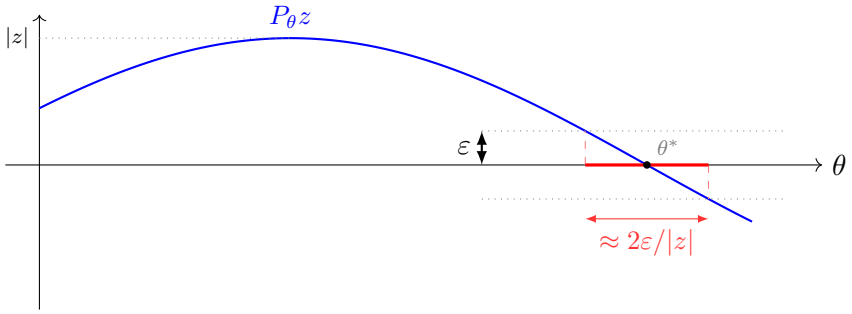


Figure 4: Transversality for $k = 1$, $N = 2$. Write $P_\theta = P_{V_\theta}$, $V_\theta = \mathbb{R}e_\theta$, identified with the map $z \mapsto \langle z, e_\theta\rangle = z_1 \cos\theta + z_2 \sin\theta$. The graph of $\theta \mapsto P_\theta z$ vanishes at the angle θ^* with $e_{\theta^*} \perp z$, and it crosses zero *transversally*: $\frac{d}{d\theta} P_\theta z = \langle z, e_{\theta+\pi/2}\rangle = \pm|z| \neq 0$ at $\theta = \theta^*$. Hence the set of bad angles, for which $|P_\theta z| \le \varepsilon$, is an interval of length $\approx 2\varepsilon/|z|$.

Fix n and $\delta > 0$. Since $\mathcal{H}^k(Z_n) \le \mathcal{H}^k(X - X) = 0$, we can cover Z_n by countably many balls $B(z_i, r_i)$ with $z_i \in Z_n$ and $\sum_i r_i^k < \delta$. If $V \in B_n$, then $P_V z = 0$ for some $z \in Z_n$, and $z \in B(z_i, r_i)$ for some i, hence

$$|P_V z_i| = |P_V(z_i - z)| \le |z_i - z| < r_i.$$

Thus $B_n \subset \bigcup_i \{V : |P_V z_i| \le r_i\}$. By Lemma 1.15 and $|z_i| \ge 2^{-n}$,

$$\gamma_{k,N}(B_n) \le \sum_i C_{N,k} \Big(\frac{r_i}{|z_i|}\Big)^k \le C_{N,k}\, 2^{nk} \sum_i r_i^k < C_{N,k}\, 2^{nk}\, \delta.$$

As $\delta > 0$ is arbitrary, $\gamma_{k,N}(B_n) = 0$, and hence $\gamma_{k,N}(B) = 0$.

It remains to check the sufficient conditions. This follows from inequalities

$$\dim_{\mathrm{H}}(X - X) \le \overline{\dim}_{\mathrm{B}}(X - X) \le \overline{\dim}_{\mathrm{B}}(X \times X) \le 2\overline{\dim}_{\mathrm{B}} X.$$

References: [SYC91, Thm. 4.6], [Rob11, Thm. 6.2], [BGŚ20, Thm. 3.7]. □

It turns out that the condition $2\dim_{\mathrm{H}} X < k$ is *not* sufficient for Theorem 1.14 to hold:

Example 1.17 ($\dim_{\mathrm{H}} X$ does not control $\dim_{\mathrm{H}}(X - X)$)**.** (i) There exist compact sets $X, Y \subset [0,1]$ with $\dim_{\mathrm{H}} X = \dim_{\mathrm{H}} Y = 0$ and $X + Y \supset [0,1]$ [Rob11, Section 2.3], [Fal14]. Hence $Z = X \cup (-Y)$ satisfies $\dim_{\mathrm{H}} Z = 0$ and $\dim_{\mathrm{H}}(Z - Z) = 1$, since $Z - Z \supset X - (-Y) = X + Y \supset [0,1]$.

(ii) This can be used to construct, for every $m \ge 2$, a compact set $K \subset \mathbb{R}^m$ with $\dim_{\mathrm{H}} K = 0$ such that no linear map $L\colon \mathbb{R}^m \to \mathbb{R}^k$ with $k < m$ is injective on K [SYC91, Appendix A], [Rob11, Section 6.1]. In particular, Theorem 1.14 is not valid under the assumption $2\dim_{\mathrm{H}} X < k$.

Remark 1.18 (History). Mañé [Mañ81] considered compact subsets of Banach spaces and topologically generic projections, and stated the result under the assumption $k > 2\dim_{\mathrm{H}} X + 1$. This is not sufficient (see [ER85, p. 627] and Example 1.17(ii)); the proof actually requires a bound on $\dim_{\mathrm{H}}(X - X)$. The version for compact subsets of Banach spaces and a prevalent set of linear maps $L\colon B \to \mathbb{R}^k$ under $\dim_{\mathrm{H}}(X - X) < k$ is [Rob11, Thm. 6.2].

Example 1.19 (The factor 2 is necessary)**.** If X is the polyhedron of Example 1.2, then $\dim_{\mathrm{H}} X = \overline{\dim}_{\mathrm{B}} X = n$ and no continuous injective map $X \to \mathbb{R}^{2n}$ exists; in particular no linear map $\mathbb{R}^N \to \mathbb{R}^{2n}$ is injective on X. Concretely: no linear map $\mathbb{R}^3 \to \mathbb{R}^2$ is injective on $K_5 \subset \mathbb{R}^3$. Thus $2\overline{\dim}_{\mathrm{B}} X < k$ in Theorem 1.14 cannot be relaxed even to $2\overline{\dim}_{\mathrm{B}} X \le k$.

1.5 Mañé's theorem for measures

Definition 1.20 (Almost surely injective maps)**.** Let μ be a Borel measure on a metric space X. A map $\phi\colon X \to \mathbb{R}^k$ is *μ-almost surely injective* if there exists a Borel set $X_\phi \subset X$ with $\mu(X \setminus X_\phi) = 0$ such that $\phi|_{X_\phi}$ is injective.

Theorem 1.21 (Probabilistic embedding theorem)**.** *Let $X \subset \mathbb{R}^N$ be a Borel set with $\mathcal{H}^k(X) = 0$ and let μ be a σ-finite Borel measure on X. Then P_V is μ-almost surely injective for $\gamma_{k,N}$-a.e. $V \in \mathrm{Gr}(k, N)$.*

Proof. This is the same as proof of Theorem 1.14, but easier. We can assume w.l.o.g. that X is σ-compact (by regularity of μ).

Step 1: bad subspaces for a fixed point. For $x \in X$ let

$$B_x = \{V \in \mathrm{Gr}(k, N) : P_V x = P_V y \text{ for some } y \in X \setminus \{x\}\}.$$

We show that

$$\gamma_{k,N}(B_x) = 0, \tag{2}$$

arguing as in the proof of Theorem 1.14, with $X-X$ replaced by $x-X$. Let $Y_n = \{y \in X : |x-y| \geq 2^{-n}\}$, so that $B_x = \bigcup_{n=1}^\infty B_{x,n}$ with $B_{x,n} = \{V \in \mathrm{Gr}(k, N) : P_V x = P_V y \text{ for some } y \in Y_n\}$. Fix n and $\delta > 0$. Since $\mathcal{H}^k(Y_n) \leq \mathcal{H}^k(X) = 0$, we can cover Y_n by countably many balls $B(y_i, r_i)$ with $y_i \in Y_n$ and $\sum_i r_i^k < \delta$. If $V \in B_{x,n}$, then $P_V x = P_V y$ for some $y \in Y_n$, and $y \in B(y_i, r_i)$ for some i, hence

$$|P_V(x - y_i)| = |P_V(y - y_i)| \leq |y - y_i| < r_i.$$

By Lemma 1.15 applied to $z = x - y_i$, $|z| \geq 2^{-n}$,

$$\gamma_{k,N}(B_{x,n}) \leq \sum_i C_{N,k}\Big(\frac{r_i}{|x - y_i|}\Big)^k \leq C_{N,k}\, 2^{nk} \sum_i r_i^k < C_{N,k}\, 2^{nk}\, \delta.$$

As $\delta > 0$ is arbitrary, $\gamma_{k,N}(B_{x,n}) = 0$, and hence $\gamma_{k,N}(B_x) = 0$.

Step 2: Fubini's theorem. Let

$$B = \{(x, V) \in X \times \mathrm{Gr}(k, N) : V \in B_x\} = \{(x, V) : P_V x = P_V y \text{ for some } y \in X \setminus \{x\}\}.$$

B is Borel (as we assume X to be σ-compact), and by Fubini's theorem and Step 1,

$$\int_{\mathrm{Gr}(k,N)} \mu(\{x \in X : (x, V) \in B\})\, d\gamma_{k,N}(V) = (\mu \otimes \gamma_{k,N})(B) = \int_X \gamma_{k,N}(B_x)\, d\mu(x) = 0.$$

Hence, for $\gamma_{k,N}$-a.e. V, the Borel set $X_V = \{x \in X : (x, V) \notin B\}$ satisfies $\mu(X \setminus X_V) = 0$. Finally, P_V is injective on X_V:

$$\text{if } x \in X_V, \text{ then } P_V x \neq P_V y \qquad \text{for every } y \in X \setminus \{x\}. \tag{3}$$

Reference: [BGŚ20, Thm. 3.1]. □

One can state the theorem in terms of the measure alone:

Corollary 1.22. *Let μ be a finite Borel measure on $\mathbb{R}^N$ with $\mu \perp \mathcal{H}^k$ (it suffices to assume that $\dim_{\mathrm{H}} \mu < k$). Then P_V is μ-a.s. injective for $\gamma_{k,N}$-a.e. $V \in \mathrm{Gr}(k, N)$.*

Proof. Choose a Borel set X of full μ-measure with $\mathcal{H}^k(X) = 0$ and apply Theorem 1.21. [BGŚ20, Cor. 3.4]. □

Corollary 1.22 was first obtained by Alberti, Bölcskei, De Lellis, Koliander and Riegler [ABD+19, Thm. II.1] under the stronger assumption $\underline{\dim}_{\mathrm{MB}} \mu < k$, where $\underline{\dim}_{\mathrm{MB}} \mu$ is the lower modified box-counting dimension of μ. One always has $\dim_{\mathrm{H}} \mu \leq \underline{\dim}_{\mathrm{MB}} \mu$, and the inequality can be strict: there are measures with $\dim_{\mathrm{H}} \mu = 1 < \underline{\dim}_{\mathrm{MB}} \mu = 2$ [BGŚ20, Thm. 5.5].

Remark 1.23 (Why the factor 2 disappears). In the proof of Theorem 1.14 one covers the *set of pairs* $X - X$ (dimension up to $2 \dim X$). In the proof of Theorem 1.21, thanks to Fubini, one point x is fixed and only X has to be covered. Note also that the relevant dimension is the Hausdorff dimension of the *measure*, not of its support (Example 1.12). For instance, the Klein bottle K (a compact surface, $\dim K = 2$) embeds into $\mathbb{R}^4$, but not into $\mathbb{R}^3$. Nevertheless, for a Klein bottle $K \subset \mathbb{R}^4$, a typical projection to $\mathbb{R}^3$ is injective outside a set of zero area, by Theorem 1.21 applied to $\mu = \mathcal{H}^2|_K$ (see Figure 5).

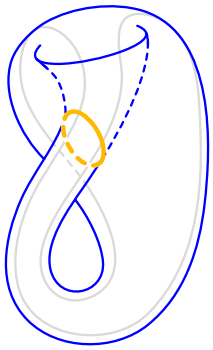


Figure 5: The Klein bottle immersed in $\mathbb{R}^3$ (a projection of a Klein bottle embedded in $\mathbb{R}^4$). The map is not injective, but it is injective outside the self-intersection circle (orange), which has zero area.

Example 1.24 (Sharpness: $\dim_{\mathrm{H}} \mu \le k$ is not enough)**.** Let σ be the normalized surface measure on the unit sphere $S^k \subset \mathbb{R}^{k+1}$ (so $\dim_{\mathrm{H}} \sigma = k$).

(i) No linear map $L\colon \mathbb{R}^{k+1} \to \mathbb{R}^k$ is σ-a.s. injective. In particular, P_V is not σ-a.s. injective for any $V \in \mathrm{Gr}(k, k+1)$.

(ii) No Lipschitz map $\phi\colon S^1 \to \mathbb{R}$ is σ-a.s. injective [BGŚ20, Example 3.5].

(iii) There exists a continuous map $\psi\colon S^1 \to \mathbb{R}$ which is σ-a.s. injective.

Proof. (i) Since $\dim \ker L \ge 1$, we can fix a unit vector $v \in \ker L$. The reflection $R_v(x) = x - 2\langle x, v\rangle v$ in the hyperplane $v^\perp$ is linear and preserves the norm, since $|v| = 1$ gives

$$|R_v x|^2 = |x|^2 - 4\langle x, v\rangle^2 + 4\langle x, v\rangle^2 |v|^2 = |x|^2.$$

Hence R_v is orthogonal, so it preserves S^k and σ. Moreover, $L \circ R_v = L$, as $Lv = 0$. Suppose that L is injective on a Borel set $A \subset S^k$ with $\sigma(A) = 1$. Then also $\sigma(R_v(A)) = 1$, and as the great sphere $\{x \in S^k : \langle x, v\rangle = 0\}$ is σ-null, there exists $x \in A \cap R_v(A)$ with $\langle x, v\rangle \ne 0$. Then $R_v x \in A$ (as R_v is an involution), $R_v x \ne x$ and $L(R_v x) = Lx$, a contradiction.

(ii) Suppose that a C-Lipschitz map $\phi\colon S^1 \to \mathbb{R}$ is injective on a Borel set $A \subset S^1$ with $\sigma(A) = 1$. Let $a = \min \phi$ and $b = \max \phi$; then $a < b$, as A has more than one point. Fix $x, y \in S^1$ with $\phi(x) = a$, $\phi(y) = b$, and let I, J be the two open arcs forming $S^1 \setminus \{x, y\}$. By the intermediate value theorem, every $c \in (a, b)$ is attained by ϕ both on I and on J, so it has at least two preimages, and at most one of them lies in A. Therefore $(a, b) \subset \phi(S^1 \setminus A)$. On the other hand, $\mathcal{H}^1(\phi(E)) \le C\,\mathcal{H}^1(E)$ for every $E \subset S^1$, so $\mathrm{Leb}_1(\phi(S^1 \setminus A)) = 0$, a contradiction.

(iii) Let $F\colon [0,1] \to [0,1]$ be an increasing homeomorphism which maps a Borel set $S \subset [0,1]$ with $\mathrm{Leb}_1(S) = 1$ onto a Lebesgue-null set $F(S)$ (e.g. the distribution function of a non-atomic, fully supported probability measure on $[0,1]$ which is singular with respect to Leb_1). Identify S^1 with $[0,1)$ via $t \mapsto e^{2\pi i t}$ (so that σ corresponds to Leb_1) and let

$$\psi(t) = \begin{cases} F(2t), & t \in [0, \frac{1}{2}], \\ 2 - 2t, & t \in [\frac{1}{2}, 1]. \end{cases}$$

Then ψ is continuous on S^1 (as $\psi(0) = \psi(1) = 0$ and $F(1) = 1$), strictly increasing on $[0, \frac{1}{2}]$ and strictly decreasing on $[\frac{1}{2}, 1]$. Let

$$A_1 = \{t \in (0, \tfrac{1}{2}) : 2t \in S\}, \qquad A_2 = \{t \in (\tfrac{1}{2}, 1) : 2 - 2t \notin F(S)\}.$$

Then $\mathrm{Leb}_1(A_1) = \mathrm{Leb}_1(A_2) = \frac{1}{2}$, while $\psi(A_1) \subset F(S)$ and $\psi(A_2) \subset [0,1] \setminus F(S)$. Hence ψ is injective on the set $A_1 \cup A_2$ of full measure. □

Remark 1.25 (Injective projections in *every* direction). There is a compactly supported Borel probability measure μ on $\mathbb{R}^2$ with $\dim_{\mathrm{H}} \mu > 0$ such that *every* orthogonal projection onto a line is μ-a.s. injective [BGŚ26b, Thm. 1.9]. This phenomenon cannot occur for sets: if $X \subset \mathbb{R}^N$ contains two points $x \neq y$, then P_V is not injective on X for every $V \in \mathrm{Gr}(k, N)$ with $x - y \in V^\perp$ (such V exist, as $k < N$).

Optimality. For sets, the condition $2\overline{\dim}_{\mathrm{B}} X < k$ in Theorem 1.14 cannot be improved in general (Example 1.19), but it need not be optimal for a given set X: a.e. projection may be injective on X also for some $k < 2\overline{\dim}_{\mathrm{B}} X$. For instance, a line segment $X \subset \mathbb{R}^2$ and the unit circle $S^1 \subset \mathbb{R}^2$ both have upper box-counting dimension 1, but for $k = 1$ the projection P_V is injective on the segment for all but one $V \in \mathrm{Gr}(1, 2)$ (cf. the proof of Example 1.13), while no P_V is injective on the circle. Thus, for $\dim X \leq k < 2 \dim X$, the answer depends on the geometry of X, not only on its dimension. For measures, the situation is cleaner: as we explain in the next section, $\dim_{\mathrm{H}} \mu$ is the exact threshold.

1.6 Connections with projection theorems and conditional measures

We relate almost sure injectivity to conditional measures and to the classical projection and slicing theorems, and prove that the Hausdorff dimension of a measure is the exact threshold for almost sure injectivity of typical projections. For a map ϕ and a measure μ, we write $\phi\mu = \mu \circ \phi^{-1}$ for the push-forward of μ under ϕ.

Definition 1.26 (Conditional measures)**.** Let $X \subset \mathbb{R}^N$ be compact, μ a Borel probability measure on X and $\phi\colon X \to \mathbb{R}^k$ continuous. A family $\{\mu_y\}_{y \in \mathbb{R}^k}$ of Borel measures on X is a *system of conditional measures* of μ with respect to ϕ if

(i) for every $y \in \mathbb{R}^k$, μ_y is a (possibly zero) Borel measure concentrated on the fibre $\phi^{-1}(y)$,

(ii) for $\phi\mu$-a.e. $y \in \mathbb{R}^k$, μ_y is a probability measure,

(iii) for every Borel set $A \subset X$, the function $y \mapsto \mu_y(A)$ is $\phi\mu$-measurable and

$$\mu(A) = \int_{\mathbb{R}^k} \mu_y(A)\, d\phi\mu(y).$$

By Rokhlin's disintegration theorem [Roh52], a system of conditional measures exists and is unique up to a $\phi\mu$-null set of y's. Moreover, by Simmons' theorem [Sim12], for $\phi\mu$-a.e. y,

$$\mu_y = \lim_{\varepsilon \to 0} \frac{\mu|_{\phi^{-1}(B(y,\varepsilon))}}{\mu(\phi^{-1}(B(y,\varepsilon)))} \qquad \text{in the weak-}* \text{ topology},$$

i.e. μ_y is the limit of the normalized restrictions of μ to the preimages of shrinking balls around y. For a compactly supported Borel probability measure μ on $\mathbb{R}^N$ and $V \in \mathrm{Gr}(k, N)$, we denote by $\{\mu_a^V\}_{a \in V}$ the system of conditional measures of μ with respect to P_V; the measure μ_a^V is concentrated on the affine plane $P_V^{-1}(a) = a + V^\perp$.

Lemma 1.27 (A.s. injectivity via conditional measures)**.** *Let $X \subset \mathbb{R}^N$ be compact, μ a Borel probability measure on X, $\phi\colon X \to \mathbb{R}^k$ continuous, and let $\{\mu_y\}_{y \in \mathbb{R}^k}$ be a system of conditional measures of μ with respect to ϕ. Then the following conditions are equivalent:*

(a) ϕ *is* μ*-a.s. injective,*

(b) $\mu_{\phi(x)} = \delta_x$ *for* μ*-a.e.* $x \in X$*,*

(c) μ_y *is a Dirac measure for* $\phi\mu$*-a.e.* $y \in \mathbb{R}^k$*.*

Proof. (a)⇒(b),(c). Assume that ϕ is injective on a Borel set $X_\phi \subset X$ with $\mu(X \setminus X_\phi) = 0$. By Definition 1.26(iii) applied to $A = X \setminus X_\phi$,

$$\int_{\mathbb{R}^k} \mu_y(X \setminus X_\phi)\, d\phi\mu(y) = \mu(X \setminus X_\phi) = 0,$$

so $\mu_y(X \setminus X_\phi) = 0$ for $\phi\mu$-a.e. y. Together with Definition 1.26(i) and (ii), this shows that for $\phi\mu$-a.e. y, the measure μ_y is a probability measure concentrated on $X_\phi \cap \phi^{-1}(y)$. As ϕ is injective on X_ϕ, this set consists of a single point x_y, and hence $\mu_y = \delta_{x_y}$. Consequently, for μ-a.e. $x \in X_\phi$ we have $\mu_{\phi(x)} = \delta_{x_{\phi(x)}} = \delta_x$, since x is the only point of $X_\phi \cap \phi^{-1}(\phi(x))$.

(b)⇒(a). Assume that $\mu_{\phi(x)} = \delta_x$ for all x in a Borel set $X' \subset X$ of full μ-measure. If $x, x' \in X'$ and $\phi(x) = \phi(x')$, then $\delta_x = \mu_{\phi(x)} = \mu_{\phi(x')} = \delta_{x'}$, so $x = x'$. Hence ϕ is injective on X'.

(c)⇒(a). Assume that $\mu_y = \delta_{s(y)}$ for y in a Borel set $G \subset \mathbb{R}^k$ with $\phi\mu(G) = 1$. By Definition 1.26(i), $\phi(s(y)) = y$ for $y \in G$; in particular, s is injective and ϕ is injective on $s(G)$. By Definition 1.26(iii), for every Borel set $A \subset X$,

$$\mu(A) = \int_G \delta_{s(y)}(A)\, d\phi\mu(y) = \phi\mu(s^{-1}(A)),$$

so $\mu = s(\phi\mu)$, and hence $\mu(s(G)) = \phi\mu(G) = 1$. Thus ϕ is injective on the set $s(G)$ of full μ-measure. (To see that $s(G)$ can be taken Borel, shrink G so that $s|_G$ is Borel and use the Lusin–Souslin theorem.)

Reference: [BGŚ26b, Lemma 2.6]. □

Projection theorems. Theorems 1.14 and 1.21 are about *injectivity* of typical projections. The classical results concern the *dimension* of typical projections and of typical slices.

Theorem 1.28 (Projection and slicing theorems)**.** *Let* $X \subset \mathbb{R}^N$ *be a Borel set and* μ *a finite Borel measure on* $\mathbb{R}^N$*. For* $\gamma_{k,N}$*-a.e.* $V \in \mathrm{Gr}(k, N)$*:*

(i) $\dim_{\mathrm{H}} P_V X = \min\{k, \dim_{\mathrm{H}} X\}$*, and* $\mathcal{H}^k(P_V X) > 0$ *if* $\dim_{\mathrm{H}} X > k$*;*

(ii) $\dim_{\mathrm{H}} P_V\mu = \min\{k, \dim_{\mathrm{H}} \mu\}$*,* $\underline{\dim}_{\mathrm{H}} P_V\mu = \min\{k, \underline{\dim}_{\mathrm{H}}\mu\}$*, and* $P_V\mu \ll \mathcal{H}^k$ *if* $\underline{\dim}_{\mathrm{H}}\mu > k$*;*

(iii) if μ *is a compactly supported probability measure with* $\underline{\dim}_{\mathrm{H}}\mu > k$*, then the conditional measures of* μ *with respect to* P_V *satisfy*

$$\underline{\dim}_{\mathrm{H}}\mu_a^V \geq \underline{\dim}_{\mathrm{H}}\mu - k \qquad \text{for } P_V\mu\text{-a.e. } a \in V.$$

Part (i) is due to Marstrand [Mar54] for $N = 2$, $k = 1$, Kaufman [Kau68] and Mattila [Mat75], see also [Mat95, Ch. 9]. Part (ii) is due to Hu and Taylor [HT94]; for the absolute continuity see [Mat95, Ch. 9]. Part (iii) is the slicing theorem of Järvenpää and Mattila [JM98, Thm. 3.3], see also [Mat95, Thm. 10.7]; it is formulated there for *sliced measures*, which for a.e. V coincide with the conditional measures μ_a^V up to normalization, see [BGŚ26b, Section 3]. For a survey of projection theorems see [FFJ15].

Example 1.29 (Injectivity does not prevent dimension drop)**.** There is no direct connection between the injectivity of a projection and the preservation of dimension. Let $f\colon [0,1] \to \mathbb{R}$ be the Weierstrass function $f(t) = \sum_{n=0}^{\infty} \lambda^n \cos(2\pi b^n t)$, where $b \geq 2$ is an integer and $1/b < \lambda < 1$.

Then f is α-Hölder with $\alpha = -\log\lambda/\log b \in (0,1)$, and its graph $\Gamma_f = \{(t, f(t)) : t \in [0,1]\}$ satisfies $\dim_{\mathrm{H}} \Gamma_f = 2 - \alpha > 1$ [She18], cf. [Fal14, Ch. 11]. The projection $P(t,y) = t$ onto the horizontal axis is injective on Γ_f, but $\dim_{\mathrm{H}} P(\Gamma_f) = \dim_{\mathrm{H}}[0,1] = 1 < \dim_{\mathrm{H}} \Gamma_f$ (see Figure 6). Conversely, the projections of the unit circle $S^1 \subset \mathbb{R}^2$ onto lines do not decrease its dimension, but none of them is injective, even almost surely (Example 1.24).

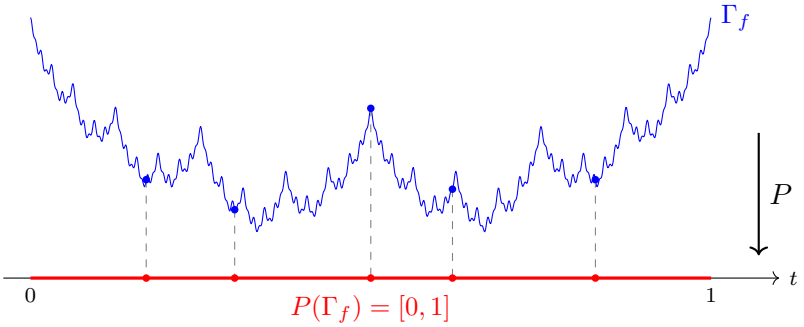


Figure 6: The graph Γ_f of the Weierstrass function with $b = 2$ and $\lambda = 0.7$ (partial sum up to $n = 7$), for which $\dim_{\mathrm{H}} \Gamma_f = 2 + \log\lambda/\log b \approx 1.49$. The projection $P(t,y) = t$ onto the horizontal axis (dashed lines) is injective on Γ_f, but its image $P(\Gamma_f) = [0,1]$ (red) has dimension 1.

Proposition 1.30 ($\dim_{\mathrm{H}} \mu$ is the threshold [BGŚ26b])**.** *Let μ be a finite Borel measure on $\mathbb{R}^N$ with $\dim_{\mathrm{H}} \mu > k$. Then for $\gamma_{k,N}$-a.e. V, P_V is not injective on any Borel set of full μ-measure.*

Proof. Fix s with $k < s < \dim_{\mathrm{H}} \mu$. By Proposition 1.11, the set $\{x \in \mathbb{R}^N : \underline{d}(\mu, x) > s\}$ has positive μ-measure, so it contains a compact set K with $\mu(K) > 0$. Let $\nu = \mu(K)^{-1}\mu|_K$. As $\nu(B(x,r)) \le \mu(K)^{-1}\mu(B(x,r))$, we have $\underline{d}(\nu, x) \ge \underline{d}(\mu, x) > s$ for $x \in K$, so $\underline{\dim}_{\mathrm{H}} \nu \ge s > k$ by Proposition 1.11. By Theorem 1.28(iii), for $\gamma_{k,N}$-a.e. V we have $\underline{\dim}_{\mathrm{H}} \nu_a^V \ge s - k > 0$ for $P_V\nu$-a.e. $a \in V$. Since $\underline{\dim}_{\mathrm{H}} \delta_x = 0$, for such V the conditional measures ν_a^V are not Dirac measures for $P_V\nu$-a.e. a, so P_V is not ν-a.s. injective by Lemma 1.27. Finally, if P_V were injective on a Borel set Y with $\mu(\mathbb{R}^N \setminus Y) = 0$, then also $\nu(\mathbb{R}^N \setminus Y) = 0$, and P_V would be ν-a.s. injective. Hence, for $\gamma_{k,N}$-a.e. V, the projection P_V is not injective on any Borel set of full μ-measure.

Reference: [BGŚ26b, Prop. 1.4]. □

Together with Corollary 1.22, Proposition 1.30 shows that for measures the Hausdorff dimension is the exact threshold: if $\dim_{\mathrm{H}} \mu < k$, then P_V is μ-a.s. injective for $\gamma_{k,N}$-a.e. V, while if $\dim_{\mathrm{H}} \mu > k$, then for $\gamma_{k,N}$-a.e. V it is not. Only in the border case $k = \dim_{\mathrm{H}} \mu$ does the answer depend on μ.

Remark 1.31 (The border case $\dim_{\mathrm{H}} \mu = k$). Both behaviours occur: $\mathcal{H}^1$ on a segment in $\mathbb{R}^2$ is projected injectively by a.e. projection onto a line, while $\mathcal{H}^1$ on a circle is not (Example 1.24). The border case is related to an open problem of Mattila on purely unrectifiable 1-sets, see [BGŚ26b, Remark 1.5].

1.7 From projections to linear and Lipschitz maps; prevalence

The results obtained so far for orthogonal projections extend almost for free to random linear maps and to typical Lipschitz maps. We explain this in the present section.

Let $\mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k) \simeq \mathbb{R}^{kN}$ denote the space of linear maps $L\colon \mathbb{R}^N \to \mathbb{R}^k$, equipped with the Lebesgue measure; "a.e. L" will always refer to this measure.

Remark 1.32 (Linear maps versus projections). Lebesgue-a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$ $(k \le N)$ has full rank and can be written uniquely as

$$L = M \circ P_V, \qquad \text{where} \quad V = (\ker L)^\perp \in \mathrm{Gr}(k, N)$$

and $M\colon V \to \mathbb{R}^k$ is a linear isomorphism [Rob11, Lemma 6.1]. Under this identification full-measure sets of linear maps correspond to full-measure sets of $V \in \mathrm{Gr}(k, N)$, and injectivity (resp. Hölder continuity of the inverse) of L and of P_V on a given set are equivalent [BGŚ26b, Remark 3.1]. As a corollary, one obtains for free the versions of the previous theorem.

The advantage of linear maps is that they form a vector space, so they can be added to a given nonlinear map $\phi\colon X \to \mathbb{R}^k$. They are also the natural model for the applications in analog compression, where a signal is compressed by linear measurements (see Section 1.8). For *linear perturbations* $\phi + L$ one needs a version of Lemma 1.15 with a translation term. Let

$$E = E(N,k) = \big\{L\colon \mathbb{R}^N \to \mathbb{R}^k :\ Lx = (\langle l_1, x\rangle, \dots, \langle l_k, x\rangle) \text{ with } l_1, \dots, l_k \in B_N(0,1)\big\},$$

equipped with the normalized Lebesgue measure γ (identify E with $B_N(0,1)^k$). By scaling, it is enough to prove "a.e. L" statements for γ-a.e. $L \in E$. Note that $\|L\| \le \sqrt{N}$ for $L \in E$.

Lemma 1.33 (Key estimate for linear maps [HK99])**.** *For every $x \in \mathbb{R}^N \setminus \{0\}$, $z \in \mathbb{R}^k$ and $\varepsilon > 0$,*

$$\gamma(\{L \in E : |Lx + z| \le \varepsilon\}) \le \Big(C\sqrt{N}\,\frac{\varepsilon}{|x|}\Big)^k,$$

where $C > 0$ is an absolute constant.

Sketch of proof. In the case $z = 0$, the proof is essentially the same as the one of Lemma 1.15. One can observe that $z = 0$ is the worst case. We leave the details to the reader.

Reference: [Rob11, Lemma 4.1]. □

Prevalence. In infinite-dimensional spaces of maps (e.g. Lipschitz maps $X \to \mathbb{R}^k$, or observables in Part 3) there is no Lebesgue measure; the following notion replaces "almost every".

Definition 1.34 (Prevalence [Chr72, HSY92])**.** Let V be a complete metrizable topological vector space. A Borel set $S \subset V$ is *prevalent* if there exists a Borel measure ν on V with $0 < \nu(K) < \infty$ for some compact $K \subset V$, such that $v + e \in S$ for ν-a.e. $e \in V$, for every $v \in V$. A non-Borel set is prevalent if it contains a prevalent Borel set. In particular, if there are $e_1, \dots, e_m \in V$ (a *probe set*) such that for every $v \in V$ we have $v + \sum_j \alpha_j e_j \in S$ for Lebesgue-a.e. $(\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$, then S is prevalent.

Proposition 1.35. *(i) In $\mathbb{R}^n$ a set is prevalent if and only if its complement has Lebesgue measure zero.*

(ii) Prevalent sets are dense; countable intersections of prevalent sets are prevalent.

(iii) In an infinite-dimensional Banach space every compact set has prevalent complement.

Reference: [HSY92], [OY05], [Rob11, Ch. 5].

Spaces of Lipschitz maps. For $X \subset \mathbb{R}^N$, let $\mathrm{Lip}(X, \mathbb{R}^k)$ be the space of Lipschitz maps $\phi\colon X \to \mathbb{R}^k$, equipped with the norm

$$\|\phi\|_{\mathrm{Lip}} = |\phi(x_0)| + \sup_{x,y \in X,\ x \ne y} \frac{|\phi(x) - \phi(y)|}{|x - y|},$$

where $x_0 \in X$ is fixed. It is a Banach space containing (the restrictions to X of) all linear maps $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$; for bounded X one can equivalently replace $|\phi(x_0)|$ by $\sup_{x\in X} |\phi(x)|$. For locally Lipschitz maps (i.e. maps $\phi\colon X \to \mathbb{R}^k$ which are Lipschitz on every bounded subset of X), the norm is replaced by the family of seminorms $p_m(\phi) = \|\phi|_{X\cap B(0,m)}\|_{\mathrm{Lip}}$, $m \in \mathbb{N}$, which generate the metric

$$d(\phi, \psi) = \sum_{m=1}^{\infty} 2^{-m} \min\{1, p_m(\phi - \psi)\}.$$

This turns the space of locally Lipschitz maps into a complete metrizable topological vector space; convergence in d means convergence in $\|\cdot\|_{\mathrm{Lip}}$ on every bounded subset of X.

Theorem 1.36 (Lipschitz perturbations)**.** *Let $X \subset \mathbb{R}^N$ and fix a (locally) Lipschitz map $\phi\colon X \to \mathbb{R}^k$.*

(i) If $\mathcal{H}^k(X \times X) = 0$ (e.g. if $\dim_{\mathrm{H}}(X \times X) < k$ or $2\overline{\dim}_{\mathrm{B}} X < k$), then $\phi + L$ is injective on X for a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$.

(ii) If X is a Borel set with $\mathcal{H}^k(X) = 0$ (e.g. $\dim_{\mathrm{H}} X < k$) and μ is a σ-finite Borel measure on X, then $\phi + L$ is μ-a.s. injective for a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$.

Consequently, both statements hold for prevalent (locally) Lipschitz maps.

Sketch of proof. (i) As in the proof of Theorem 1.14, with $X - X$ replaced by the set $Z = \{(x - y, \phi(x) - \phi(y)) : x, y \in X\} \subset \mathbb{R}^N \times \mathbb{R}^k$, which is the image of $X \times X$ under a (locally) Lipschitz map, so $\mathcal{H}^k(Z) = 0$. Cover $\{(z_1, z_2) \in Z : |z_1| \geq 2^{-n}\}$ by balls $B((a_i, b_i), r_i)$ with $\sum_i r_i^k < \delta$; if $Lz_1 + z_2 = 0$ for some $(z_1, z_2) \in B((a_i, b_i), r_i)$, then $|La_i + b_i| \leq (\|L\| + 1)r_i$, and apply Lemma 1.33 to $x = a_i$ and $z = b_i$; cf. [Rob11, Lemma 14.4]. (ii) As in the proof of Theorem 1.21, applying Lemma 1.33 to $x - y_i$ and $z = \phi(x) - \phi(y_i)$; [BGŚ20, Thm. 3.1]. □

1.8 Why bother? On applications

Compressed sensing. The classical theory of compressed sensing [CRT06, Don06, FR13] shows that a vector $x \in \mathbb{R}^N$ with at most s non-zero coordinates can be recovered from $k \approx s \log(N/s)$ random linear measurements $Lx \in \mathbb{R}^k$. Its strength is that it provides an effective decoding algorithm (ℓ^1-minimization), which works with high probability. From the point of view of these notes, the set X of such vectors is a finite union of s-dimensional coordinate subspaces, so $X - X$ is contained in a finite union of $2s$-dimensional subspaces, and a.e. linear map $L\colon \mathbb{R}^N \to \mathbb{R}^k$ with $k > 2s$ is injective on X by Theorem 1.14 and Remark 1.32. An important and actively studied task is to remove the sparsity assumption and instead put constraints on the geometric complexity of the signal, modelled as a random vector or a stochastic process [WV10, ABD+19, SRAB17, BGŚ26b, JP17, GŚ26].

Analog compression of stochastic processes. In (almost) lossless analog compression, a signal modelled by a stochastic process $(X_i)_{i=1}^{\infty}$ is compressed by linear measurements of $(X_1, \ldots, X_n)$, and one asks for the minimal measurements rate which allows (almost) lossless recovery. This can be seen as a non-discrete version of Shannon's compression theory for discrete sources, with the entropy replaced by a suitable notion of (mean) dimension of the source. Some of the main tools come from geometric measure theory and are similar to the ones we have seen so far and will see further. Wu and Verdú [WV10] initiated the theory for memoryless sources, where the fundamental limits are expressed in terms of the Rényi information dimension. Jalali and Poor [JP17] showed that for ψ^*-mixing processes, a rate of random linear measurements exceeding the mean information dimension of the process suffices for almost lossless recovery (see also [RJEP17, GK19]), and [GŚ26] proved, using geometric measure theory methods, that this rate is optimal. For further works on analog compression of stochastic processes see also [WV12, JMB14, ZBD15, RJEP17, GK19, GŚ19, GŚ20].

Regularity of the decompressor. If we treat applying a linear map or an orthogonal projection as a compression, then its inverse on the (full-measure) set of injectivity is the decompression. In practice the measurements are perturbed by noise, so the decompressor should be stable: its regularity (continuity, Hölder or Lipschitz continuity) governs the quality of the compression. For instance, if g is α-Hölder with constant C and the measurement Lx is known up to an error ε, then x is recovered up to an error $C\varepsilon^{\alpha}$. Therefore, we will study next the regularity of the inverse.

Part 2: Regularity of the embedding

2.1 Regularity of the embedding: definitions

If X is a compact metric space and $\phi\colon X \to \mathbb{R}^k$ is continuous and injective, then $\phi^{-1}\colon \phi(X) \to X$ is automatically continuous. The question is whether it has a quantitative modulus of continuity. The answer depends on the notion of dimension (cf. the table at the end of Section 2.2).

Definition 2.1 (Hölder inverses)**.** Let (X, ρ) be a metric space, $\phi\colon X \to \mathbb{R}^k$ a Lipschitz map and $0 < \alpha \le 1$.

(i) ϕ has an *α-Hölder inverse* on X if there is $C > 0$ with $\rho(x,y) \le C|\phi(x) - \phi(y)|^{\alpha}$ for all $x, y \in X$. If ϕ has a 1-Hölder (Lipschitz) inverse, then ϕ is a *bi-Lipschitz embedding*.

(ii) ϕ has a *pointwise α-Hölder inverse at $x \in X$* if there is $C_x > 0$ with $\rho(x,y) \le C_x|\phi(x) - \phi(y)|^{\alpha}$ for all $y \in X$.

(iii) Let μ be a finite Borel measure on X. The map ϕ is *μ-nearly bi-Lipschitz* if for μ-a.e. x it has a pointwise α-Hölder inverse at x for every $\alpha \in (0,1)$ [BSŚ24, Def. 2.6].

Note that regularity of the inverse includes injectivity: if ϕ has a pointwise α-Hölder inverse at x, then $\phi(y) \neq \phi(x)$ for all $y \neq x$. Similarly a map with a pointwise α-Hölder inverse at every point of a set $X' \subset X$ is injective on X'.

For measures one cannot expect a uniformly Hölder inverse (Remark 2.5), and the pointwise notions in (ii) and (iii) are the right ones.

2.2 The regularity theorem

As in Section 1.7, we consider perturbations $\phi_L = \phi + L$ of a fixed Lipschitz map ϕ by a typical linear map L. For sets we obtain uniform regularity of the inverse under conditions involving $2\dim$, while for measures we obtain pointwise regularity under conditions involving $\dim$ — the same gain of the factor 2 as in Part 1.

Theorem 2.2 (Regularity of the embedding)**.** *Let $X \subset \mathbb{R}^N$ be compact and fix a Lipschitz map $\phi\colon X \to \mathbb{R}^k$. For $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$ write $\phi_L = \phi + L$.*

(A) Sets: uniform regularity.

(i) If $\dim_{\mathrm{H}}(X \times X) < k$ (for $\phi = 0$ it suffices that $\dim_{\mathrm{H}}(X - X) < k$), then for a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$ the map ϕ_L is injective on X, so it has a continuous inverse $(\phi_L)^{-1}\colon \phi_L(X) \to X$.

(ii) If $2\overline{\dim}_{\mathrm{B}} X < k$, then for a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$ and every $\alpha \in \big(0, 1 - \frac{2\overline{\dim}_{\mathrm{B}} X}{k}\big)$, the map ϕ_L has an α-Hölder inverse on X.

(iii) If $\phi = 0$ and $\dim_{\mathrm{A}}(X - X) < k$, then for a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$ and every $\alpha \in (0,1)$, the map L has an α-Hölder inverse on X.

(B) Measures: pointwise regularity.

Let μ be a finite Borel measure on X. Then for a.e. $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$:

- *(i) if $\dim_{\mathrm{H}} \mu < k$, then ϕ_L is μ-a.s. injective; if moreover $\dim_{\mathrm{H}} X < k$, then for μ-a.e. $x \in X$ the inverse is continuous at $\phi_L(x)$, i.e. for every $\varepsilon > 0$ there is $\delta > 0$ such that $|\phi_L(x) - \phi_L(y)| \le \delta$ implies $|x - y| \le \varepsilon$ for all $y \in X$;*
- *(ii) if $\overline{\dim}_{\mathrm{B}} X < k$, then for μ-a.e. $x \in X$ the map ϕ_L has a pointwise α-Hölder inverse at x, for every $\alpha \in (0, 1 - \frac{\overline{\dim}_{\mathrm{B}} X}{k})$;*
- *(iii) if $\dim_{\mathrm{A}} X < k$, then ϕ_L is μ-nearly bi-Lipschitz.*

The same statements hold for the orthogonal projections P_V, $\gamma_{k,N}$-a.e. $V \in \mathrm{Gr}(k, N)$. As in Theorem 1.36, (A)(i), (A)(ii) and (B) are prevalence statements in $\mathrm{Lip}(X, \mathbb{R}^k)$.

Reference: (A)(ii) [BAEFN93], [HK99], [Rob11, Thm. 4.3]; (A)(iii) [Ols02], [Rob11, Thm. 9.18]; (B) [BGŚ26b, Thms. 1.7, 1.12, 4.2, 4.4]. The classical formulation of (A)(ii) concerns linear maps; its proof gives the stated version for Lipschitz perturbations as well, cf. [Rob11, Thm. 4.3]. We prove (B) in Section 2.3, based on [BGŚ26b].

Remark 2.3. The assumption in (A)(iii) cannot be replaced by $2\dim_{\mathrm{A}} X < k$ or $\dim_{\mathrm{A}}(X \times X) < k$: although $X - X$ is the image of $X \times X$ under the Lipschitz map $(x, y) \mapsto x - y$, the Assouad dimension may increase under Lipschitz maps. For a nonlinear Lipschitz map ϕ, the same argument shows that if

$$\dim_{\mathrm{A}} Z_\phi < k, \qquad \text{where} \quad Z_\phi = \{(x - y, \phi(x) - \phi(y)) : x, y \in X\} \subset \mathbb{R}^N \times \mathbb{R}^k,$$

then for a.e. L the map ϕ_L has an α-Hölder inverse on X for every $\alpha \in (0, 1)$. For linear ϕ, the set Z_ϕ is a bi-Lipschitz image of $X - X$, so $\dim_{\mathrm{A}} Z_\phi = \dim_{\mathrm{A}}(X - X)$.

Remark 2.4 (Logarithmic bounds). In (A)(iii) and (B)(iii) one has in fact more precise, *almost bi-Lipschitz* (up to a logarithmic factor) bounds. Let $R \ge \operatorname{diam} X$. In (A)(iii), for every $\gamma > \frac{1}{k - \dim_{\mathrm{A}}(X - X)}$ there is $C > 0$ with

$$|\phi_L(x) - \phi_L(y)| \ge C \frac{|x - y|}{\log(2R/|x - y|)^\gamma} \qquad \text{for all } x, y \in X$$

[Ols02], [Rob11, Thm. 9.18]. In (B)(iii), for every $\gamma > \frac{1}{k - \dim_{\mathrm{A}} X}$ and μ-a.e. x there is $C_x > 0$ with

$$|\phi_L(x) - \phi_L(y)| \ge C_x \frac{|x - y|}{\log(2R/|x - y|)^\gamma} \qquad \text{for all } y \in X$$

[BGŚ26b]. The proofs require a finer, logarithmic choice of scales.

Remark 2.5 (No uniform Hölder continuity for measures). In (B) one cannot ask for a uniformly Hölder inverse. Let $X \subset \mathbb{R}^N$ be a compact set with $\dim_{\mathrm{A}} X < k$ which does not embed topologically into $\mathbb{R}^k$ (e.g. the set from Example 1.2 with $n < k \le 2n$), and let $\operatorname{supp} \mu = X$. Then no continuous $\phi \colon X \to \mathbb{R}^k$ is injective on a set of full μ-measure with a (uniformly) Hölder inverse: the inverse would extend continuously (in fact: Hölder continuously) to $\overline{\phi(X_\phi)}$, giving a topological embedding of $X = \overline{X_\phi}$ into $\mathbb{R}^k$. [BGŚ20, Remark 3.10], [BGŚ26b, Remark 1.8].

Remark 2.6 (Sharpness). The assumptions in Theorem 2.2(B) are sharp in the following sense [BGŚ26b, Section 6]: in the second part of (B)(i), $\dim_{\mathrm{H}} X$ cannot be replaced by $\dim_{\mathrm{H}} \mu$; in (B)(ii), $\overline{\dim}_{\mathrm{B}} X$ cannot be replaced by $\dim_{\mathrm{H}} X$, and the range of exponents is optimal; in (B)(iii), $\dim_{\mathrm{A}} X$ cannot be replaced by $\overline{\dim}_{\mathrm{B}} X$. Moreover, in (B)(iii) the map ϕ_L need not be bi-Lipschitz on any set of full μ-measure, even if $\dim_{\mathrm{A}} X = 0$: there is a compact set $X \subset \mathbb{R}^2$ with $\dim_{\mathrm{A}} X = 0$ and $\dim_{\mathrm{A}} P_V X = 1$ for every $V \in \mathrm{Gr}(1, 2)$ [FK20, Thm. 2.1], so if $\operatorname{supp} \mu = X$, then no linear map $L \colon \mathbb{R}^2 \to \mathbb{R}$ is bi-Lipschitz on a set of full μ-measure.

Summary of Parts 1 and 2.

object	assumption	conclusion for a.e. L	reference
set X	$\dim_{\mathrm{H}}(X-X) < k$	continuous inverse	Thm. 2.2(A)(i)
set X	$2\overline{\dim}_{\mathrm{B}} X < k$	α-Hölder inverse, $\alpha < 1 - \frac{2\overline{\dim}_{\mathrm{B}} X}{k}$	Thm. 2.2(A)(ii)
set X	$\dim_{\mathrm{A}}(X-X) < k$	α-Hölder inverse, $\alpha < 1$	Thm. 2.2(A)(iii)
measure μ	$\dim_{\mathrm{H}} \mu > k$	not a.s. injective	Prop. 1.30
measure μ	$\dim_{\mathrm{H}} \mu < k$	a.s. injective	Thm. 2.2(B)(i)
measure μ	$\dim_{\mathrm{H}} \operatorname{supp} \mu < k$	a.s. injective, continuous inverse	Thm. 2.2(B)(i)
measure μ	$\overline{\dim}_{\mathrm{B}} \operatorname{supp} \mu < k$	pointwise α-Hölder, $\alpha < 1 - \frac{\overline{\dim}_{\mathrm{B}} \operatorname{supp} \mu}{k}$	Thm. 2.2(B)(ii)
measure μ	$\dim_{\mathrm{A}} \operatorname{supp} \mu < k$	nearly bi-Lipschitz	Thm. 2.2(B)(iii)

2.3 Proof of the regularity theorem for measures

We prove part (B) of Theorem 2.2. We work with perturbations $\phi_L = \phi + L$ by linear maps $L \in \mathrm{Lin}(\mathbb{R}^N, \mathbb{R}^k)$; for $\phi = 0$, by Remark 1.32, this is the same as working with orthogonal projections. As in Section 1.7, by scaling it suffices to prove the assertions for γ-a.e. $L \in E$; recall that $\|L\| \le \sqrt{N}$ for $L \in E$. We denote by $\mathrm{Lip}(\phi)$ the Lipschitz constant of ϕ.

Sketch of proof of (B)(i). The injectivity part is Theorem 1.36(ii), applied to a Borel set of full μ-measure with Hausdorff dimension smaller than k. For the continuity part, note that $\mathcal{H}^k(X) = 0$ as $\dim_{\mathrm{H}} X < k$. Hence the proof of Theorem 1.36(ii), applied to the compact set X, shows that for γ-a.e. $L \in E$ and μ-a.e. $x \in X$ we have $\phi_L(x) \neq \phi_L(y)$ for every $y \in X \setminus \{x\}$ (this is the same as (3) in the case of orthogonal projections). Fix such L and x, and let $\varepsilon > 0$. The continuous function $y \mapsto |\phi_L(x) - \phi_L(y)|$ is positive on the compact set $\{y \in X : |x-y| \ge \varepsilon\}$, so it has a positive minimum 2δ there. Hence $|\phi_L(x) - \phi_L(y)| \le \delta$ implies $|x-y| < \varepsilon$. See also [BGŚ26b]. □

The proof of (B)(ii) and (B)(iii) is based on the following estimates of the size of the "bad" set.

Lemma 2.7 (Covering estimate)**.** *Let $X \subset \mathbb{R}^N$ be compact and $x \in X$. For $0 < \varepsilon < \delta/2$ let*

$$B_x(\varepsilon, \delta) = \{L \in E : |\phi_L(x) - \phi_L(y)| \le \varepsilon \text{ for some } y \in X \text{ with } |x-y| \ge \delta\}.$$

For every $\theta > 0$ there exists $D = D(X, \mathrm{Lip}(\phi), \theta) > 0$, independent of x, ε and δ, such that

(a) $\gamma(B_x(\varepsilon,\delta)) \le D\, \varepsilon^{k - \overline{\dim}_{\mathrm{B}} X - \theta}\, \delta^{-k}$;

(b) if $\dim_{\mathrm{A}} X + \theta < k$, *then* $\gamma(B_x(\varepsilon,\delta)) \le D\, (\varepsilon/\delta)^{k - \dim_{\mathrm{A}} X - \theta}$.

Proof. We may assume $\delta \le \operatorname{diam} X$, as otherwise $B_x(\varepsilon,\delta) = \emptyset$. Let $K = 1 + \sqrt{N} + \mathrm{Lip}(\phi)$. If $L \in B_x(\varepsilon,\delta)$ is witnessed by $y \in X$ and $y \in B(w,\varepsilon)$ for some $w \in X$, then

$$|x-w| \ge |x-y| - \varepsilon \tag{4}$$

and, as $\|L\| \le \sqrt{N}$,

$$|L(x-w) + \phi(x) - \phi(w)| = |\phi_L(x) - \phi_L(w)| \le |\phi_L(x) - \phi_L(y)| + (\|L\| + \mathrm{Lip}(\phi))|y-w| \le K\varepsilon. \tag{5}$$

(a) By the definition of $\overline{\dim}_{\mathrm{B}} X$, the set X can be covered by m balls $B(y_j, \varepsilon)$ with $y_j \in X$, where

$$m \le D_\theta\, \varepsilon^{-\overline{\dim}_{\mathrm{B}} X - \theta}$$

and D_θ depends only on X and θ (recall that $\varepsilon < \operatorname{diam} X$). As $\delta - \varepsilon \geq \delta/2$, this gives by (4) and (5)

$$B_x(\varepsilon,\delta) \subset \bigcup_{j=1}^{m} \{L \in E : |L(x-y_j)+\phi(x)-\phi(y_j)| \leq K\varepsilon \text{ and } |x-y_j| \geq \delta/2\}.$$

By Lemma 1.33 applied to $x - y_j$ and $z = \phi(x) - \phi(y_j)$, each set in the union has γ-measure at most $(2C\sqrt{N}K\varepsilon/\delta)^k$ (it might be empty if $|x - y_j| < \delta/2$). Hence

$$\gamma(B_x(\varepsilon,\delta)) \leq D_\theta\, \varepsilon^{-\overline{\dim}_{\mathrm{B}} X-\theta}\Big(\frac{2C\sqrt{N}K\varepsilon}{\delta}\Big)^k.$$

(b) Let $s = \dim_{\mathrm{A}} X + \theta < k$. We have $\{y \in X : |x-y| \geq \delta\} = \bigcup_{n=0}^{\infty} X_n$, where $X_n = \{y \in X : 2^n\delta \leq |x-y| < 2^{n+1}\delta\}$. As $X_n \subset B(x, 2^{n+1}\delta)$, the definition of the Assouad dimension gives a cover of X_n by m_n balls $B(y_{j,n}, \varepsilon)$, $j = 1, \ldots, m_n$, with $y_{j,n} \in X$, where

$$m_n \leq M\Big(\frac{2^{n+1}\delta}{\varepsilon}\Big)^s.$$

As $2^n\delta - \varepsilon \geq 2^{n-1}\delta$, we obtain as in (a) by (4) and (5)

$$B_x(\varepsilon,\delta) \subset \bigcup_{n=0}^{\infty}\bigcup_{j=1}^{m_n} \{L \in E : |L(x-y_{j,n})+\phi(x)-\phi(y_{j,n})| \leq K\varepsilon \text{ and } |x-y_{j,n}| \geq 2^{n-1}\delta\}.$$

By Lemma 1.33, each set in the union has γ-measure at most $\big(C\sqrt{N}K\varepsilon/(2^{n-1}\delta)\big)^k$. Hence

$$\gamma(B_x(\varepsilon,\delta)) \leq \sum_{n=0}^{\infty} m_n\Big(\frac{C\sqrt{N}K\varepsilon}{2^{n-1}\delta}\Big)^k \leq M\,2^s(2C\sqrt{N}K)^k\Big(\frac{\varepsilon}{\delta}\Big)^{k-s}\sum_{n=0}^{\infty} 2^{-n(k-s)},$$

and the series converges as $s < k$. □

Proof of Theorem 2.2(B)(ii), (iii). By scaling, we may assume that $\operatorname{diam} X \leq 1$. Let $\alpha \in (0,1)$, and in case (ii) assume moreover that $\alpha < 1 - \overline{\dim}_{\mathrm{B}} X/k$. Fix $x \in X$ and for $M > 4$ let

$$B_{x,M} = \{L \in E : |x-y| > M|\phi_L(x) - \phi_L(y)|^\alpha \text{ for some } y \in X\}.$$

Then ϕ_L has a pointwise α-Hölder inverse at x if and only if $L \notin \bigcap_M B_{x,M}$. We show that $\gamma(B_{x,M}) \to 0$ as $M \to \infty$, which gives

$$\gamma\Big(\bigcap_M B_{x,M}\Big) = 0. \tag{6}$$

In both cases $\overline{\dim}_{\mathrm{B}} X < k$, so $\mathcal{H}^k(X) = 0$, and by the proof of Theorem 1.36(ii) (Step 1 of the proof of Theorem 1.21, with Lemma 1.33 in place of Lemma 1.15), for γ-a.e. $L \in E$ we have $\phi_L(y) \neq \phi_L(x)$ for all $y \in X \setminus \{x\}$ (this is the same as (2) in the case of orthogonal projections). For such $L \in B_{x,M}$, the point y in the definition of $B_{x,M}$ satisfies $0 < |\phi_L(x) - \phi_L(y)| < M^{-1/\alpha} < 1$ (as $|x-y| \leq \operatorname{diam} X \leq 1$), so

$$2^{-(n+1)} < |\phi_L(x) - \phi_L(y)| \leq 2^{-n} \qquad \text{and} \qquad |x-y| > M2^{-\alpha(n+1)} \qquad \text{for some } n \geq 0.$$

Hence, up to a γ-null set,

$$B_{x,M} \subset \bigcup_{n=0}^{\infty} B_x(2^{-n}, M2^{-\alpha(n+1)}),$$

with the sets $B_x(\varepsilon,\delta)$ of Lemma 2.7; note that $2^{-n} < M2^{-\alpha(n+1)}/2$, as $M > 4$ and $\alpha < 1$, so Lemma 2.7 can be applied.

In case (ii), choose $\theta > 0$ with $\alpha < 1 - (\overline{\dim}_{\mathrm{B}} X + \theta)/k$. By Lemma 2.7(a),

$$\gamma(B_{x,M}) \le \sum_{n=0}^{\infty} D\, 2^{-n(k-\overline{\dim}_{\mathrm{B}} X-\theta)} M^{-k} 2^{\alpha k(n+1)} = D\, 2^{\alpha k} M^{-k} \sum_{n=0}^{\infty} 2^{-n(k-\overline{\dim}_{\mathrm{B}} X-\theta-\alpha k)},$$

and the series converges by the choice of θ. In case (iii), choose $\theta > 0$ with $s = \dim_{\mathrm{A}} X + \theta < k$. By Lemma 2.7(b),

$$\gamma(B_{x,M}) \le \sum_{n=0}^{\infty} D\Big(\frac{2^{-n}}{M 2^{-\alpha(n+1)}}\Big)^{k-s} = D\, 2^{\alpha(k-s)} M^{-(k-s)} \sum_{n=0}^{\infty} 2^{-n(1-\alpha)(k-s)},$$

and the series converges as $\alpha < 1$. In both cases $\gamma(B_{x,M}) \to 0$ as $M \to \infty$, which proves (6). Consequently, for every $x \in X$, the map ϕ_L has a pointwise α-Hölder inverse at x for γ-a.e. $L \in E$.

Let $B = \{(x, L) \in X \times E : \phi_L \text{ has no pointwise } \alpha\text{-Hölder inverse at } x\}$. As $B = \bigcap_{M\in\mathbb{N}}\{(x, L) : |x - y| > M|\phi_L(x) - \phi_L(y)|^{\alpha} \text{ for some } y \in X\}$ is an intersection of open sets, it is Borel, and by Fubini's theorem

$$\int_E \mu(\{x : (x, L) \in B\})\, d\gamma(L) = \int_X \gamma(\{L : (x, L) \in B\})\, d\mu(x) = \int_X \gamma\Big(\bigcap_M B_{x,M}\Big)\, d\mu(x) \overset{(6)}{=} 0.$$

Hence, for γ-a.e. $L \in E$, the map ϕ_L has a pointwise α-Hölder inverse at μ-a.e. x; in particular, it is injective on the set X_L of such points, which has full μ-measure. As a pointwise α-Hölder inverse is also a pointwise α'-Hölder inverse for $\alpha' < \alpha$ (recall $\operatorname{diam} X \le 1$), applying this to a sequence of exponents increasing to $1 - \overline{\dim}_{\mathrm{B}} X/k$ in case (ii), resp. to 1 in case (iii), we obtain the assertion for all exponents simultaneously.

For $\phi = 0$, the statements for orthogonal projections follow from Remark 1.32. See also [BGŚ26b, Thms. 1.7, 1.12, 4.2, 4.4]. □

The following is a useful property of μ-nearly bi-Lipschitz maps. They preserve the local dimension of *every* measure ν at μ-a.e. point.

Lemma 2.8 (Nearly bi-Lipschitz maps preserve local dimensions [BSŚ24])**.** *Let (X, ρ) be a metric space, μ a finite Borel measure on X and $\phi\colon X \to \mathbb{R}^k$ a Lipschitz map. If ϕ is μ-nearly bi-Lipschitz, then for μ-a.e. $x \in X$ and* every *finite Borel measure ν on X,*

$$\underline{d}(\phi\nu, \phi(x)) = \underline{d}(\nu, x) \qquad \textit{and} \qquad \overline{d}(\phi\nu, \phi(x)) = \overline{d}(\nu, x).$$

Proof. Fix $x \in X$ such that ϕ has a pointwise α-Hölder inverse at x for every $\alpha \in (0, 1)$; this holds for μ-a.e. x. Let ν be a finite Borel measure on X and $r > 0$. As $\phi(B(x, r/\operatorname{Lip}(\phi))) \subset B(\phi(x), r)$, we have

$$\phi\nu(B(\phi(x), r)) = \nu(\phi^{-1}(B(\phi(x), r))) \ge \nu(B(x, r/\operatorname{Lip}(\phi))).$$

On the other hand, fix $\alpha \in (0, 1)$ and $C_x > 0$ with $\rho(x, y) \le C_x|\phi(x) - \phi(y)|^{\alpha}$ for all $y \in X$. Then $\phi^{-1}(B(\phi(x), r)) \subset B(x, C_x r^{\alpha})$, so

$$\phi\nu(B(\phi(x), r)) \le \nu(B(x, C_x r^{\alpha})).$$

Taking logarithms, dividing by $\log r < 0$ and using $\frac{\log(r/\operatorname{Lip}(\phi))}{\log r} \to 1$ and $\frac{\log(C_x r^{\alpha})}{\log r} \to \alpha$ as $r \to 0$, we obtain

$$\alpha\, \underline{d}(\nu, x) \le \underline{d}(\phi\nu, \phi(x)) \le \underline{d}(\nu, x) \qquad \text{and} \qquad \alpha\, \overline{d}(\nu, x) \le \overline{d}(\phi\nu, \phi(x)) \le \overline{d}(\nu, x).$$

Letting $\alpha \to 1$ finishes the proof.

Reference: [BSŚ24, Lemma 8.1]. □

This lemma will be the key to the multifractal application in Part 4.

2.4 Bi-Lipschitz embeddings and the Assouad dimension

The strongest form of regularity is a Lipschitz inverse. It is clear that if a metric space X admits a bi-Lipschitz embedding into a Euclidean space $\mathbb{R}^n$, then it has finite Assouad dimension (in fact $\dim_{\mathrm{A}} X \le n$, as the Assouad dimension is invariant under bi-Lipschitz maps). The converse is not true: the Heisenberg group, equipped with the Carnot–Carathéodory metric [Sem96], and the Laakso–Lang–Plaut graphs [Laa02, LP01] are metric spaces of finite Assouad dimension with no bi-Lipschitz embedding into any Euclidean space (for the second one: into any Hilbert space), see [Rob11, Section 9.4]. In fact, an intrinsic characterization of metric spaces which admit a bi-Lipschitz embedding into some Euclidean space is an open problem [Hei03, Section 1.4], [NN12, Section 1]; see [DPT25] for the current state of the art. What one can always achieve for spaces of finite Assouad dimension is a bi-Lipschitz embedding after "snowflaking" the metric.

Theorem 2.9 (Assouad 1983)**.** *If (X,ρ) has finite Assouad dimension, then for every $\varepsilon \in (0,1)$ there are $n \in \mathbb{N}$ and a bi-Lipschitz embedding of the "snowflake" (X,ρ^ε) into $\mathbb{R}^n$, i.e. a map $\phi\colon X \to \mathbb{R}^n$ with $c^{-1}\rho(x,y)^\varepsilon \le |\phi(x)-\phi(y)| \le c\,\rho(x,y)^\varepsilon$ for some $c>0$.*

Reference: [Ass83], [Hei01], [Rob11, Section 9.4]; for the dimension n see [NN12].

Part 3: Dynamical systems — reconstruction and prediction from observed data

3.1 Time-delay embeddings in time series analysis

Let $T\colon X \to X$ be a (discrete-time) dynamical system, $X \subset \mathbb{R}^N$; think of the states of a physical system, with T describing the evolution of the state over a unit of time. In applications the state $x \in X$ is usually not accessible: the phase space may be unknown and high-dimensional, and one cannot measure all the coordinates of the state. Instead, one measures a single scalar quantity along an orbit: given an *observable* $h\colon X \to \mathbb{R}$, an experimentalist has access only to the *time series*

$$y_n = h(x_n), \qquad \text{where } x_n = T^n x, \quad n = 0,1,2,\dots$$

The question is what can be learned about (X,T) from the time series alone, and how to predict its future values.

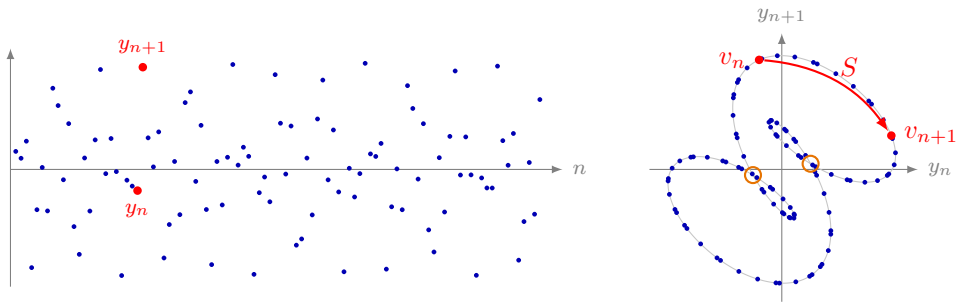
A single value $y_n = h(x_n)$ does not determine the state x_n, as h is far from being injective. The *method of delays* [PCFS80, Tak81] uses several consecutive values instead: fix $k \in \mathbb{N}$ (the *delay length*) and replace the unknown state x_n by the *delay vector*

$$v_n = (y_n, y_{n+1}, \dots, y_{n+k-1}) \in \mathbb{R}^k,$$

obtained by sliding a window of length k along the time series. The idea is to model the orbit $x_0, x_1, x_2, \dots$ of the system by the sequence $v_0, v_1, v_2, \dots$ of points in $\mathbb{R}^k$, and the dynamics $x_n \mapsto x_{n+1} = Tx_n$ by the transition $v_n \mapsto v_{n+1}$, which shifts the window by one step, i.e. drops y_n and appends the next value y_{n+k}:

$$(y_n, \dots, y_{n+k-1}) \longmapsto (y_{n+1}, \dots, y_{n+k}),$$

see Figure 7. Note that $v_n = \phi_{h,k}(x_n)$ for the delay coordinate map $\phi_{h,k}$ defined below, so the delay vectors lie in $\phi_{h,k}(X) \subset \mathbb{R}^k$. For the model to make sense, the transition should be given by a well-defined map S on $\phi_{h,k}(X)$: the next delay vector should be determined by the current one, regardless of the orbit along which it is observed. This is guaranteed if $\phi_{h,k}$ is injective, as then the delay vector determines the state, and $(\phi_{h,k}(X), S)$ is an exact copy of (X,T) in $\mathbb{R}^k$ (Proposition 3.2). Such a model is built from the data alone, which makes it useful for prediction.



(a) time series $y_n = h(T^n x_0)$ (b) delay vectors $v_n = (y_n, y_{n+1})$

Figure 7: The method of delays for the rotation T of the circle X by the angle 43°, the observable $h(x_1, x_2) = 4x_1^3 - 2x_1 + 0.35x_2$ and $x_0 = (1, 0)$. (a) The time series y_n, $n = 1, \dots, 100$. The values alone ($k = 1$) do not determine the dynamics: states with different futures give the same observation. (b) The delay vectors $v_n = (y_n, y_{n+1})$, $n = 1, \dots, 99$, with the transition $S\colon v_n \mapsto v_{n+1}$ (red; $n = 24$). The vectors lie on the closed curve $\phi_{h,2}(X)$ (grey). It has two self-intersection points (circled), where $\phi_{h,2}$ is not injective, and S is well defined everywhere else. Thus $k = 2$ delays give a faithful model of the system on a set of full measure, but not on the whole of X (cf. Theorems 3.4 and 3.8).

The delay-coordinate method is commonly used in practice, see e.g. [Aba96, KS04, Sau06] and the references therein. We will say a little bit more about that later. There is a mathematical theory aiming at explaining why and when these methods work. The main object of study in this theory are the delay-coordinate maps.

Definition 3.1 (Delay coordinate map)**.** The *k-delay coordinate map* is

$$\phi_{h,k}\colon X \to \mathbb{R}^k, \qquad \phi_{h,k}(x) = \big(h(x), h(Tx), \dots, h(T^{k-1}x)\big).$$

It is a *k-delay embedding* if it is injective. The observable h is *(deterministically) k-predictable* if for every $x_1, x_2 \in X$, $\phi_{h,k}(x_1) = \phi_{h,k}(x_2)$ implies $\phi_{h,k}(Tx_1) = \phi_{h,k}(Tx_2)$.

The following simple proposition explains the relationship between the properties of the delay-coordinate maps and the dynamics reconstructed in $\mathbb{R}^k$.

Proposition 3.2. *h is deterministically k-predictable if and only if there is a map $S\colon \phi_{h,k}(X) \to \phi_{h,k}(X)$ (the* prediction map*) with $\phi_{h,k} \circ T = S \circ \phi_{h,k}$; then $S(h(x), \dots, h(T^{k-1}x)) = (h(Tx), \dots, h(T^k x))$ and the following diagram commutes:*

$$\begin{array}{ccc} X & \xrightarrow{\;T\;} & X \\ {\scriptstyle \phi_{h,k}}\downarrow & & \downarrow{\scriptstyle \phi_{h,k}} \\ \phi_{h,k}(X) & \xrightarrow{\;S\;} & \phi_{h,k}(X) \end{array}$$

If X is compact and T, h are continuous, S is continuous. If $\phi_{h,k}$ is injective, then $S = \phi_{h,k} \circ T \circ \phi_{h,k}^{-1}$ and $(\phi_{h,k}(X), S)$ is conjugated to (X, T) via $\phi_{h,k}$.

The proof is left to the reader, cf. [BGŚ24, Prop. 1.12].

3.2 Takens-type embedding theorems

The first and classical result is Takens' theorem. It can be seen as a dynamical version of Whitney's embedding theorem (Theorem 1.3): a generic map $M \to \mathbb{R}^{2m+1}$ is an embedding even if one restricts to the delay coordinate maps.

Theorem 3.3 (Takens 1981)**.** *Let M be a compact C^2 manifold of dimension m. For pairs (T, h) of a C^2-diffeomorphism $T\colon M \to M$ and a C^2 function $h\colon M \to \mathbb{R}$, it is a generic property (open and dense in the C^2 topology) that $\phi_{h,2m+1}\colon M \to \mathbb{R}^{2m+1}$ is a C^2 embedding.*

Reference: [Tak81, Thm. 1]; expositions [Huk06], [SYC91].

Takens' theorem became commonly invoked as a mathematical justification for the validity of time-delay based techniques of modelling dynamics from observed time series, see e.g. [SM90, HGLS05, MGNS18].

Note that Takens' theorem is a genericity statement, and this is the natural kind of result in this context. In applications, the system and the observable (hence the data) are given in advance, and the experimenter has no control over them. The existence of a good observable is therefore of little use; the best one can hope for is that the given observable is not atypical. This is why we are interested in properties of *typical* observables (cf. Section 1.1).

Just as Takens' theorem is a dynamical version of Whitney's theorem, many of the results of Parts 1 and 2 can be turned into embedding results for delay coordinate maps: the family of linear maps L (or projections P_V) is replaced by the family of delay coordinate maps $\phi_{h,k}$ of observables h. For instance, another commonly invoked result is the theorem of Sauer, Yorke and Casdagli [SYC91], a dynamical version of Mañé's theorem (Theorem 1.14), which allows for applications to fractal attractors. A new feature, compared to Part 1, is an assumption on the periodic points of T.

Periodic points are an obstruction. Let $\mathrm{Per}_p(T) = \{x \in X : T^p x = x\}$. If $x \in \mathrm{Per}_p(T)$ and $p < k$, then the sequence $h(x), h(Tx), h(T^2x), \dots$ is p-periodic, so $\phi_{h,k}$ is injective on $\mathrm{Per}_p(T)$ if and only if $\phi_{h,p}$ is. For instance, if T is the identity on the circle $X = S^1$, then $\phi_{h,k}(x) = (h(x), \dots, h(x))$ is not injective for any continuous h and any k. Thus one has to assume that the sets $\mathrm{Per}_p(T)$, $p < k$, are small compared to p, and in view of Theorem 1.14, the natural condition is $2\overline{\dim}_{\mathrm{B}}\mathrm{Per}_p(T) < p$. In Takens' theorem this is taken care of by perturbing also T: for a generic diffeomorphism T the sets $\mathrm{Per}_p(T)$ are finite (Kupka–Smale theorem [PdM82, Ch. 3, Thm. 3.6]). In the proofs, periodic points manifest themselves as a drop of the rank of the relevant transversality matrix (Lemma 3.11 below).

We use the Lipschitz formulation of [Rob11, Thm. 14.5] as stated in [BGŚ20, Thm. 1.1].

Theorem 3.4 (Sauer–Yorke–Casdagli)**.** *Let $X \subset \mathbb{R}^N$ be compact and $T\colon X \to X$ Lipschitz and injective. Let $k > 2\overline{\dim}_{\mathrm{B}}X$ and assume $2\overline{\dim}_{\mathrm{B}}\mathrm{Per}_p(T) < p$ for $p = 1, \dots, k-1$. Then for a prevalent Lipschitz observable $h \in \mathrm{Lip}(X, \mathbb{R})$, the delay coordinate map $\phi_{h,k}$ is injective on X.*

Reference: [SYC91, Thm. 4.13], [Rob11, Thm. 14.5], [BGŚ20, Thm. 1.1] (with the probe set consisting of polynomials of degree at most $2k$).

Remark 3.5 (Other Takens-type theorems). Infinite-dimensional systems [Rob05], [Rob11, Thm. 14.7]; continuous observables and covering dimension [Cab00, Gut16, GQS18]; forced and stochastic systems [Sta99, SBDH97]; non-invertible maps [Tak02]. The full delay map $x \mapsto (h(T^n x))_{n\in\mathbb{Z}}$ is an equivariant embedding into the shift on $\mathbb{R}^{\mathbb{Z}}$; embeddings of $\mathbb{Z}$-actions into shifts are governed by the *mean dimension* [Lin99, LT14, GT20].

3.3 The probabilistic approach: injectivity and predictability

For us, the interesting fact is that it has been observed numerically that one often seems to obtain a reliable reconstruction of the system with k much smaller than $2 \dim X$, see e.g. [SSOY98] (Hénon and Ikeda maps) and [Sau06] (the Lorenz attractor, reconstructed with $k = 3$). In 1998, Schroer, Sauer, Ott and Yorke [SSOY98] formulated conjectures concerning this phenomenon in a probabilistic setting, in which the initial point is chosen at random according to a natural measure μ of the system: then $k > \dim \mu$ delays should suffice for *prediction*. The probabilistic

point of view turned out to successfully address these issues. The notion of predictability used in [SSOY98] is the following.

Definition 3.6 (Prediction error [SSOY98])**.** Let μ be a Borel probability measure on X and $\phi = \phi_{h,k}$. For $y \in \mathbb{R}^k$ such that $\mu(\phi^{-1}(B(y,\varepsilon))) > 0$ for every $\varepsilon > 0$, and for $\varepsilon > 0$, let $\mu_{y,\varepsilon}$ be the normalized restriction of μ to $\phi^{-1}(B(y,\varepsilon))$, i.e.

$$\mu_{y,\varepsilon}(A) = \frac{\mu(A \cap \phi^{-1}(B(y,\varepsilon)))}{\mu(\phi^{-1}(B(y,\varepsilon)))} \qquad \text{for Borel } A \subset X,$$

and let

$$m_\varepsilon(y) = \int \phi \circ T \, d\mu_{y,\varepsilon}, \qquad \sigma_\varepsilon(y) = \Big(\int |\phi \circ T - m_\varepsilon(y)|^2 \, d\mu_{y,\varepsilon} \Big)^{1/2}, \qquad \sigma(y) = \lim_{\varepsilon \to 0} \sigma_\varepsilon(y).$$

A point y is *k-predictable* if $\sigma(y) = 0$; h is *almost surely k-predictable* if $\phi(x)$ is k-predictable for μ-a.e. $x \in X$.

The quantities $m_\varepsilon(y)$ and $\sigma_\varepsilon(y)$ have a simple probabilistic meaning, see Figure 8. They also govern the behaviour of the so-called Farmer–Sidorowich prediction algorithm [FS87], see more in Section 3.5.

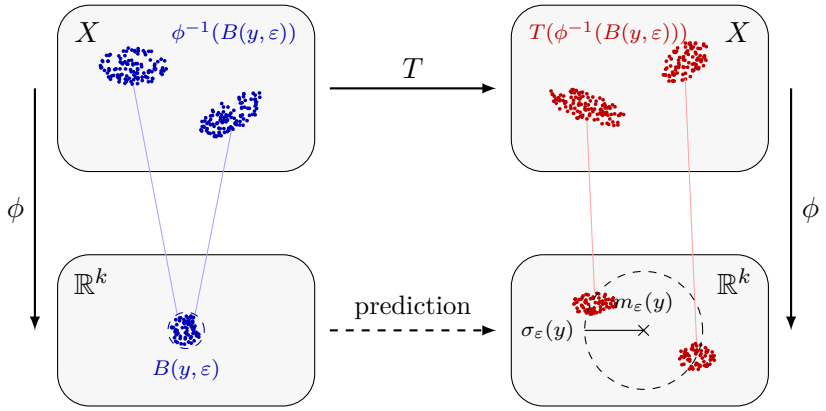


Figure 8: The meaning of Definition 3.6. Knowing the past up to accuracy ε means knowing that the delay vector $\phi(x)$ lies in a small ball $B(y,\varepsilon)$ in the model space $\mathbb{R}^k$ (bottom left). Its preimage $\phi^{-1}(B(y,\varepsilon))$ in the phase space X can be large (here it has two components, as ϕ need not be injective), and $\mu_{y,\varepsilon}$ is μ conditioned on this preimage. The dynamics T acts on X, and the result is observed through ϕ (right). The clouds of points represent probability distributions: of $x \sim \mu_{y,\varepsilon}$ and of $\phi(x)$ (blue), and of Tx and $\phi(Tx)$ (red). Then $m_\varepsilon(y)$ and $\sigma_\varepsilon(y)$ are the mean and the standard deviation of the random vector $\phi(Tx)$, $x \sim \mu_{y,\varepsilon}$: the prediction of the next delay vector and the prediction error. The point y is predictable if $\sigma_\varepsilon(y) \to 0$ as $\varepsilon \to 0$, i.e. if the dashed arrow becomes a well-defined prediction map, cf. Proposition 3.2.

Almost sure predictability has the following equivalent, deterministic description (cf. Proposition 3.2). We postpone its proof to Section 3.4.

Proposition 3.7. *Let $X \subset \mathbb{R}^N$ be compact, μ a Borel probability measure on X, let $T\colon X \to X$ and $h\colon X \to \mathbb{R}$ be continuous and $\phi = \phi_{h,k}$. Then h is almost surely k-predictable if and only if there is a Borel set $X_h \subset X$ of full μ-measure such that*

$$\phi(x_1) = \phi(x_2) \implies \phi(Tx_1) = \phi(Tx_2) \qquad \text{for } x_1, x_2 \in X_h, \tag{7}$$

i.e. there is a prediction map $S\colon \phi(X_h) \to \mathbb{R}^k$ with $\phi \circ T = S \circ \phi$ on X_h. In particular, if ϕ is μ-a.s. injective, then h is almost surely k-predictable.

Theorem 3.8 (Probabilistic Takens theorem [BGŚ20, BGŚ22, BGŚ24, BGŚ26a])**.** *Let $X \subset \mathbb{R}^N$ be compact, μ a Borel probability measure on X, $T\colon X \to X$ Lipschitz and $k \in \mathbb{N}$. Then for a prevalent Lipschitz observable $h \in \mathrm{Lip}(X, \mathbb{R})$, with $\phi = \phi_{h,k}$:*

(i) (almost sure injectivity) *if $k > \dim_{\mathrm{H}} \mu$, T is injective and $\dim_{\mathrm{H}}(\mu|_{\mathrm{Per}_p(T)}) < p$ for $p = 1, \dots, k-1$, then ϕ is μ-a.s. injective;*

(ii) (almost sure predictability) *if $k > \dim_{\mathrm{H}} \mu$, then h is almost surely k-predictable;*

(iii) (prediction error) *if $k > \overline{\dim}_{\mathrm{B}} X$, then for every $\delta, \eta > 0$ there is $C > 0$ such that for small $\varepsilon > 0$,*

$$\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \le C\varepsilon^{k - \overline{\dim}_{\mathrm{B}} X - \eta};$$

(iv) (too few delays) *if μ is T-invariant and ergodic and $k < \dim_{\mathrm{H}} \mu$, then there is $C > 0$ such that for every sufficiently small $\delta > 0$ and $0 < \varepsilon < \varepsilon_0(\delta)$,*

$$\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \ge C.$$

Reference: (i): [BGŚ20, Thm. 4.3]; (ii): [BGŚ22, Thm. 1.7] under the assumptions of (i), [BGŚ24, Thm. 1.5] in general; (iii): [BGŚ24, Thm. 1.7]; (iv): [BGŚ26a, Thms. 1.14, 1.16]. For (i) and (ii), the probe set can be taken to be a basis of the space of polynomials on $\mathbb{R}^N$ of degree at most $2k-1$.

Theorem 3.8 directly answers the conjectures of Schroer, Sauer, Ott and Yorke, after necessary changes. The interested reader can find the precise statements of the conjectures and their relation to Theorem 3.8 in Section 3.4.

Compared to Theorem 3.4, the number of delays is halved and the box-counting dimension of X is replaced by the Hausdorff dimension of μ. For instance, for the Lorenz attractor ($\dim_{\mathrm{H}} \approx 2.06$), $k = 3$ delays suffice almost surely, while Theorem 3.4 requires $k \ge 5$. We prove (i), and deduce from it (ii) under the assumptions of (i). The proof of (ii) in general requires an additional combinatorial argument, which we only state (see the end of this subsection). We do not prove (iii) and (iv).

The tools. As in Part 1, the proof of Theorem 3.8(i) is based on transversality, but the tools have to be modified. Fix a Lipschitz observable h and let $h_1, \dots, h_m$ be a basis of the space of polynomials on $\mathbb{R}^N$ of degree at most $2k-1$ (the probe set). For $\alpha \in \mathbb{R}^m$ put $h_\alpha = h + \sum_j \alpha_j h_j$. By Definition 1.34, it suffices to prove that the assertion of Theorem 3.8(i) holds for h_α, for Leb_m-a.e. $\alpha \in B_m(0,1)$. The role of the perturbation $\phi + L$ by a linear map is played by the perturbation h_α of the observable, and the perturbed delay map is *affine* in the parameter:

$$\phi_{h_\alpha,k}(x) - \phi_{h_\alpha,k}(y) = D_{x,y}\,\alpha + w_{x,y}, \tag{8}$$

where $w_{x,y} = \phi_{h,k}(x) - \phi_{h,k}(y)$ and $D_{x,y}$ is the $k \times m$ matrix

$$D_{x,y} = \begin{pmatrix} h_1(x) - h_1(y) & h_2(x) - h_2(y) & \cdots & h_m(x) - h_m(y) \\ h_1(Tx) - h_1(Ty) & h_2(Tx) - h_2(Ty) & \cdots & h_m(Tx) - h_m(Ty) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(T^{k-1}x) - h_1(T^{k-1}y) & h_2(T^{k-1}x) - h_2(T^{k-1}y) & \cdots & h_m(T^{k-1}x) - h_m(T^{k-1}y) \end{pmatrix},$$

i.e.

$$D_{x,y} = \left[h_j(T^i x) - h_j(T^i y)\right]_{0 \le i < k,\ 1 \le j \le m}.$$

Transversality is now governed by the *rank* (and the singular values) of the $k \times m$ matrix $D_{x,y}$: Lemma 3.9 below replaces Lemma 1.15, and Lemma 3.11 shows that the rank drops at periodic points.

Recall that every linear map $D\colon \mathbb{R}^m \to \mathbb{R}^k$ has a *singular value decomposition* $D = U\Sigma W^T$, where $U \in O(k)$, $W \in O(m)$ and Σ is a $k \times m$ diagonal matrix with diagonal entries $\sigma_1(D) \geq \sigma_2(D) \geq \cdots \geq \sigma_{\min\{k,m\}}(D) \geq 0$, the *singular values* of D. Equivalently, $\sigma_1(D)^2 \geq \cdots \geq \sigma_{\min\{k,m\}}(D)^2$ are the largest $\min\{k,m\}$ eigenvalues of $D^T D$. Writing w_i and u_i for the columns of W and U, we have $Dw_i = \sigma_i(D)u_i$, so D maps the unit ball $B_m(0,1)$ onto a (possibly degenerate) ellipsoid with semi-axes $\sigma_i(D)$ in the directions u_i. In particular, $\sigma_1(D) = \|D\|$, and $\sigma_p(D) > 0$ if and only if $\operatorname{rank} D \geq p$.

Lemma 3.9 (Singular values [SYC91])**.** *Let $D\colon \mathbb{R}^m \to \mathbb{R}^k$ be linear with p-th largest singular value $\sigma_p(D) > 0$. Then for $z \in \mathbb{R}^k$ and $\varepsilon > 0$,*

$$\operatorname{Leb}_m(\{\alpha \in B_m(0,1) : |D\alpha + z| \leq \varepsilon\}) \leq 2^m \Big(\frac{\varepsilon}{\sigma_p(D)}\Big)^p.$$

The lemma and its proof are illustrated in Figure 9.

Proof. Let $D = U\Sigma W^T$ be the singular value decomposition. For $\alpha \in \mathbb{R}^m$ put $\beta = W^T\alpha$ and $z' = U^T z$. Then $|D\alpha + z| = |\Sigma\beta + z'| \geq |\sigma_i(D)\beta_i + z_i'|$ for every i. Hence the isometry W^T maps the set $\{\alpha \in B_m(0,1) : |D\alpha + z| \leq \varepsilon\}$ into

$$\Big\{\beta \in [-1,1]^m : \Big|\beta_i + \frac{z_i'}{\sigma_i(D)}\Big| \leq \frac{\varepsilon}{\sigma_i(D)} \quad \text{for } i = 1,\dots,p\Big\},$$

whose Lebesgue measure is at most $\prod_{i=1}^p \frac{2\varepsilon}{\sigma_i(D)} \cdot 2^{m-p} \leq 2^m (\frac{\varepsilon}{\sigma_p(D)})^p$. □

Reference: [SYC91, Lemma 4.2], [Rob11, Lemma 14.3].

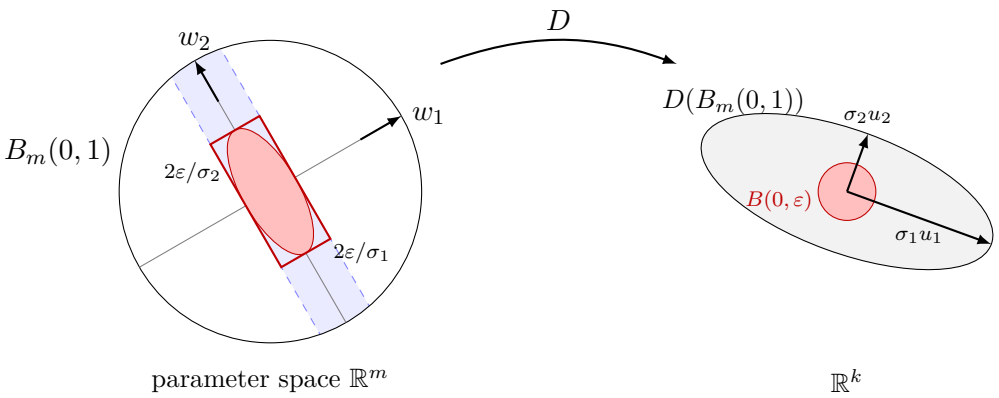


Figure 9: The proof of Lemma 3.9 for $m = k = 2$ and $z = 0$. The map D sends the unit ball $B_m(0,1)$ onto an ellipse with semi-axes $\sigma_1 = \sigma_1(D)$ and $\sigma_2 = \sigma_2(D)$ in the directions u_1, u_2, with $Dw_i = \sigma_i u_i$. Hence the set $\{\alpha : |D\alpha| \leq \varepsilon\} = D^{-1}(B(0,\varepsilon))$ (red) is an ellipse with semi-axes ε/σ_1 and ε/σ_2 in the directions w_1, w_2, contained in a rectangle with sides $2\varepsilon/\sigma_1$ and $2\varepsilon/\sigma_2$. For $p = 1$ one uses only the strip of width $2\varepsilon/\sigma_1$ in the direction w_1 (shaded). For general z, the red set is translated.

Lemma 3.10 (Interpolation)**.** *Let $h_1, \dots, h_m$ be a basis of the space of polynomials on $\mathbb{R}^N$ of degree at most $k-1$. Then for every $s \leq k$ and all distinct points $x_1, \dots, x_s \in \mathbb{R}^N$, the $s \times m$ matrix $[h_j(x_i)]_{1\leq i\leq s,\ 1\leq j\leq m}$ has rank s.*

Proof. The crucial fact is that polynomials of degree at most $k-1$ are *interpolating*: for all distinct points $x_1, \dots, x_s \in \mathbb{R}^N$ with $s \leq k$ and all values $y_1, \dots, y_s \in \mathbb{R}$, there is a polynomial P

of degree at most $k-1$ with $P(x_i) = y_i$ for $i = 1, \dots, s$. Indeed, choose $v \in \mathbb{R}^N$ such that the numbers $t_i = \langle x_i, v\rangle$ are distinct, and take the Lagrange interpolation polynomial

$$P(x) = \sum_{l=1}^{s} y_l \prod_{i \neq l} \frac{\langle x, v\rangle - t_i}{t_l - t_i},$$

which has degree $s - 1 \leq k - 1$.

Now, the matrix $A = [h_j(x_i)]$ defines the linear map $\mathbb{R}^m \to \mathbb{R}^s$,

$$c \mapsto Ac = \Big(\sum_{j=1}^{m} c_j h_j(x_i)\Big)_{i=1}^{s},$$

which assigns to c the vector of values of the polynomial $\sum_j c_j h_j$ at the points $x_1, \dots, x_s$. As $h_1, \dots, h_m$ is a basis, every polynomial of degree at most $k-1$ is of this form, so by the interpolation property the map A is onto $\mathbb{R}^s$, hence $\operatorname{rank} A = s$. □

Lemma 3.11 (Rank lemma [BGŚ20])**.** *Let T be injective and let $h_1, \dots, h_m$ be a basis of the space of polynomials on $\mathbb{R}^N$ of degree at most $2k-1$. Let $\operatorname{Orb}(x) = \{y \in X : T^n y = x \text{ or } T^n x = y \text{ for some } n \geq 0\}$ denote the full orbit of x. Then:*

(i) if $y \neq x$, then $\operatorname{rank} D_{x,y} \geq 1$;

(ii) if $y \notin \operatorname{Orb}(x)$ and y has minimal period $p \in \{1, \dots, k-1\}$, then $\operatorname{rank} D_{x,y} \geq p$;

(iii) if $y \notin \operatorname{Orb}(x)$ and y is not periodic with minimal period smaller than k, then $\operatorname{rank} D_{x,y} = k$.

Proof. Let $z_1, \dots, z_s$ be the distinct points among $x, Tx, \dots, T^{k-1}x, y, Ty, \dots, T^{k-1}y$, preserving the order. Then $s \leq 2k$, so the $s \times m$ matrix $V = [h_j(z_l)]$ has rank s by Lemma 3.10. For $0 \leq i < k$ let $a(i), b(i)$ be the indices with $z_{a(i)} = T^i x$ and $z_{b(i)} = T^i y$, and let J be the $k \times s$ matrix with entries

$$J_{il} = \mathbb{1}_{\{T^i x\}}(z_l) - \mathbb{1}_{\{T^i y\}}(z_l), \qquad 0 \leq i < k,\ 1 \leq l \leq s,$$

i.e. the i-th row of J has 1 in the column $a(i)$, -1 in the column $b(i)$ and zeros elsewhere (it is zero if $a(i) = b(i)$). Then $D_{x,y} = JV$ (see Example 3.12 below), and since $V \colon \mathbb{R}^m \to \mathbb{R}^s$ is onto, $\operatorname{rank} D_{x,y} = \operatorname{rank} J$.

(i) If $y \neq x$, then $a(0) \neq b(0)$, so the row of J corresponding to $i = 0$ is non-zero.

(ii), (iii) If $y \notin \operatorname{Orb}(x)$, then the points $T^i x$ and $T^j y$ are distinct for all i, j (if $T^i x = T^j y$, then $y \in \operatorname{Orb}(x)$ by the injectivity of T). If y has minimal period $p < k$, then the indices $b(0), \dots, b(p-1)$ are distinct and different from all $a(i)$, so the $p \times p$ submatrix of J formed by the rows $i = 0, \dots, p-1$ and the columns $b(0), \dots, b(p-1)$ equals $-\mathrm{Id}$. Hence $\operatorname{rank} J \geq p$. If y is not periodic with minimal period smaller than k, then the points $y, Ty, \dots, T^{k-1}y$ are distinct (by the injectivity of T), and the same argument with all k rows gives $\operatorname{rank} J = k$. □

Reference: [BGŚ20, Lemma 4.5]; see also [Rob11, proof of Thm. 14.5].

Example 3.12 (The decomposition $D_{x,y} = JV$ for $k = 2$)**.** (a) If the points x, Tx, y, Ty are distinct, then $s = 4$, $(z_1, z_2, z_3, z_4) = (x, Tx, y, Ty)$ and

$$D_{x,y} = \begin{pmatrix} h_1(x) - h_1(y) & \cdots & h_m(x) - h_m(y) \\ h_1(Tx) - h_1(Ty) & \cdots & h_m(Tx) - h_m(Ty) \end{pmatrix}$$

$$= \underbrace{\begin{array}{c} \begin{matrix} x & Tx & y & Ty \end{matrix} \\ \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \end{array}}_{J} \underbrace{\begin{pmatrix} h_1(x) & \cdots & h_m(x) \\ h_1(Tx) & \cdots & h_m(Tx) \\ h_1(y) & \cdots & h_m(y) \\ h_1(Ty) & \cdots & h_m(Ty) \end{pmatrix}}_{V}.$$

The last two columns of J form $-\mathrm{Id}$, so $\operatorname{rank} D_{x,y} = \operatorname{rank} J = 2 = k$.

(b) If $x \neq y$ are fixed points of T, then $s = 2$, $(z_1, z_2) = (x, y)$ and

$$D_{x,y} = \begin{pmatrix} h_1(x) - h_1(y) & \cdots & h_m(x) - h_m(y) \\ h_1(x) - h_1(y) & \cdots & h_m(x) - h_m(y) \end{pmatrix}$$
$$= \underbrace{\begin{matrix} x & y \\ \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \end{matrix}}_{J} \underbrace{\begin{pmatrix} h_1(x) & \cdots & h_m(x) \\ h_1(y) & \cdots & h_m(y) \end{pmatrix}}_{V}.$$

The two rows of $D_{x,y}$ coincide, so $\operatorname{rank} D_{x,y} = \operatorname{rank} J = 1 < k$: the rank drops at periodic points, as in Lemma 3.11(ii) with $p = 1$.

Proof of Theorem 3.8(i). Write $\phi_\alpha = \phi_{h_\alpha,k}$. Let L be a common Lipschitz constant of the maps ϕ_α, $\alpha \in B_m(0,1)$ (it exists, as T, h and $h_1, \dots, h_m$ are Lipschitz on the compact set X).

As $\dim_{\mathrm{H}} \mu < k$ and $\dim_{\mathrm{H}}(\mu|_{\mathrm{Per}_p(T)}) < p$, we can choose Borel sets $E_1, \dots, E_k$ such that

$$E_p \subset \mathrm{Per}_p(T), \quad \mu(\mathrm{Per}_p(T) \setminus E_p) = 0, \quad \mathcal{H}^p(E_p) = 0 \qquad \text{for } p = 1, \dots, k-1,$$
$$E_k \subset X, \quad \mu(X \setminus E_k) = 0, \quad \mathcal{H}^k(E_k) = 0.$$

Let $Y = E_k \setminus \bigcup_{p<k}(\mathrm{Per}_p(T) \setminus E_p)$. Then $\mu(X \setminus Y) = 0$ and $Y \cap \mathrm{Per}_p(T) \subset E_p$ for $p < k$. By inner regularity, we may assume that Y is σ-compact.

Fix $x \in Y$ and let

$$B_x = \{\alpha \in B_m(0,1) : \phi_\alpha(x) = \phi_\alpha(y) \text{ for some } y \in Y \setminus \{x\}\}.$$

Our goal is to show that

$$\mathrm{Leb}_m(B_x) = 0 \qquad \text{for every } x \in Y. \tag{9}$$

As in the proof of Theorem 1.21, this suffices by Fubini's theorem. Note that by (8), $\phi_\alpha(x) = \phi_\alpha(y)$ if and only if $D_{x,y}\alpha + w_{x,y} = 0$. Split $Y \setminus \{x\}$ into two parts.

(a) $y \in \mathrm{Orb}(x) \setminus \{x\}$. By Lemma 3.11(i), $D_{x,y} \neq 0$, so $\{\alpha : D_{x,y}\alpha + w_{x,y} = 0\}$ is contained in a proper affine subspace of $\mathbb{R}^m$. As $\mathrm{Orb}(x)$ is countable, the set of such α for some $y \in \mathrm{Orb}(x) \setminus \{x\}$ is Leb_m-null.

(b) $y \in Y \setminus \mathrm{Orb}(x)$. Let $p \in \{1, \dots, k\}$ be the minimal period of y if y is periodic with minimal period smaller than k, and $p = k$ otherwise. Then $y \in E_p$, and $\sigma_p(D_{x,y}) > 0$ by Lemma 3.11(ii) (for $p < k$) or (iii) (for $p = k$). Hence

$$Y \setminus \mathrm{Orb}(x) \subset \bigcup_{p=1}^{k} \bigcup_{n=1}^{\infty} Y_{p,n}, \qquad \text{where} \quad Y_{p,n} = \{y \in E_p \setminus \mathrm{Orb}(x) : \ \sigma_p(D_{x,y}) \geq 1/n\}.$$

Therefore, it suffices to prove that $\mathrm{Leb}_m(B_{x,p,n}) = 0$ for all $p \in \{1, \dots, k\}$ and $n \in \mathbb{N}$, where

$$B_{x,p,n} = \{\alpha \in B_m(0,1) : \phi_\alpha(x) = \phi_\alpha(y) \text{ for some } y \in Y_{p,n}\}.$$

Fix p, n and $\delta > 0$. As $\mathcal{H}^p(Y_{p,n}) \leq \mathcal{H}^p(E_p) = 0$, we can cover $Y_{p,n}$ by balls $B(y_i, r_i)$ with $y_i \in Y_{p,n}$ and $\sum_i r_i^p < \delta$. If $\alpha \in B_m(0,1)$ and $\phi_\alpha(x) = \phi_\alpha(y)$ for some $y \in Y_{p,n} \cap B(y_i, r_i)$, then

$$|D_{x,y_i}\alpha + w_{x,y_i}| = |\phi_\alpha(x) - \phi_\alpha(y_i)| = |\phi_\alpha(y) - \phi_\alpha(y_i)| \leq L r_i.$$

Hence, by Lemma 3.9 and $\sigma_p(D_{x,y_i}) \geq 1/n$,

$$\mathrm{Leb}_m(B_{x,p,n}) \leq \sum_i 2^m (nLr_i)^p < 2^m (nL)^p \, \delta.$$

As $\delta > 0$ is arbitrary, $\mathrm{Leb}_m(B_{x,p,n}) = 0$.

Hence (9) holds, which completes the proof. □

Reference: [BGŚ20, Thm. 4.3].

From injectivity to predictability. By Proposition 3.7, almost sure injectivity implies almost sure predictability. Hence Theorem 3.8(i) gives Theorem 3.8(ii) under the assumptions of (i).

On the proof of (ii) in general. Without the assumptions on periodic points, $\phi_{h,k}$ need not be μ-a.s. injective; e.g. if T is the identity on S^1 and μ is the normalized length on S^1, then no $\phi_{h,k}(x) = (h(x), \dots, h(x))$ with Lipschitz h is μ-a.s. injective (Example 1.24(ii)), but every h trivially satisfies (7). In general, the proof of (ii) is based on the following combinatorial fact. It is the heart of the proof, and requires a much more refined analysis than the one in Lemma 3.11. It shows that the pairs of points with rank $D_{x,y} < k$ (e.g. pairs of periodic points) are harmless for prediction.

Proposition 3.13 (Predictability–transversality dichotomy [BGŚ24])**.** *Let $T\colon X \to X$ be an arbitrary map and let $h_1, \dots, h_m$ be a basis of the space of polynomials on $\mathbb{R}^N$ of degree at most $2k-1$. If $x, y \in X$ and $\operatorname{rank} D_{x,y} < k$, then*

$$|\phi_{h_\alpha,k}(Tx) - \phi_{h_\alpha,k}(Ty)| \le 2k\,|\phi_{h_\alpha,k}(x) - \phi_{h_\alpha,k}(y)| \qquad \text{for every } \alpha \in \mathbb{R}^m.$$

Reference: [BGŚ24, Prop. 3.1, Cor. 3.2].

Hence, if $\phi_\alpha(x) = \phi_\alpha(y)$ and $\phi_\alpha(Tx) \neq \phi_\alpha(Ty)$, then $\operatorname{rank} D_{x,y} = k$, i.e. $\sigma_k(D_{x,y}) > 0$. For such pairs one can repeat the argument of part (b) of the proof of (i) with $p = k$, followed by Fubini's theorem; this uses only $\mathcal{H}^k(E_k) = 0$ (and neither the injectivity of T nor the assumptions on periodic points). This gives (7) for ϕ_α on a set of full μ-measure, for a.e. $\alpha \in B_m(0,1)$, and Proposition 3.7 gives (ii).

3.4 The Schroer–Sauer–Ott–Yorke conjectures

We state the conjectures of [SSOY98] precisely and compare them with Theorem 3.8. They are formulated for natural measures and the information dimension.

Definition 3.14 (Natural measure, information dimension)**.** Let T be a diffeomorphism of a compact Riemannian manifold with an attractor Λ with basin $B(\Lambda)$. An invariant probability measure μ on Λ is *natural* if $\frac{1}{n}\sum_{i<n} \delta_{T^i x} \to \mu$ weakly for Lebesgue-a.e. $x \in B(\Lambda)$. The lower and upper *information dimensions* of μ are

$$\underline{\mathrm{ID}}(\mu) = \liminf_{\varepsilon\to 0} \int \frac{\log \mu(B(x,\varepsilon))}{\log \varepsilon}\, d\mu(x) \qquad \text{and} \qquad \overline{\mathrm{ID}}(\mu) = \limsup_{\varepsilon\to 0} \int \frac{\log \mu(B(x,\varepsilon))}{\log \varepsilon}\, d\mu(x).$$

If they are equal, their common value is denoted by $\mathrm{ID}(\mu)$.

SSOY Conjecture 1 (Predictability conjecture [SSOY98])**.** *Let μ be a natural measure for a smooth diffeomorphism with $\mathrm{ID}(\mu) = D$. If $k > D$, then a generic observable h is almost surely k-predictable.*

SSOY Conjecture 2 (Prediction error conjecture [SSOY98])**.** *In the setting of SSOY Conjecture 1, for generic h, fixed $\delta > 0$ and small ε:*

(i) if $k < D$, then $\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \ge C > 0$;

(ii) if $D < k < 2D$ and ϕ is not injective, then $C_1\varepsilon^{k-D} \le \mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \le C_2\varepsilon^{k-D}$;

(iii) if $D < k < 2D$ and ϕ is injective, or if $k > 2D$, then $\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) = 0$.

We now prove Proposition 3.7, which relates the prediction error to deterministic prediction. The link is provided by the conditional measures (Section 1.6).

Proof of Proposition 3.7. Let $\nu = \phi\mu$, so that h is almost surely k-predictable if and only if ν-a.e. $y \in \mathbb{R}^k$ is k-predictable. Let $\{\mu_y\}$ be the system of conditional measures of μ with respect to ϕ (Definition 1.26). The map $F = \phi \circ T \colon X \to \mathbb{R}^k$ is continuous, hence bounded. By Simmons' theorem (Section 1.6), there is a Borel set $G \subset \mathbb{R}^k$ with $\nu(G) = 1$ such that for $y \in G$, μ_y is a probability measure concentrated on $\phi^{-1}(y)$ and $\mu_{y,\varepsilon} \to \mu_y$ weakly as $\varepsilon \to 0$. For $y \in G$, applying the weak convergence to the coordinates of F and to $|F|^2$, we obtain

$$m_\varepsilon(y) = \int F \, d\mu_{y,\varepsilon} \longrightarrow \int F \, d\mu_y =: m(y) \qquad \text{and} \qquad \int |F|^2 \, d\mu_{y,\varepsilon} \longrightarrow \int |F|^2 \, d\mu_y.$$

Expanding the square, $\sigma_\varepsilon(y)^2 = \int |F|^2 \, d\mu_{y,\varepsilon} - |m_\varepsilon(y)|^2$, and hence

$$\sigma(y)^2 = \lim_{\varepsilon \to 0} \sigma_\varepsilon(y)^2 = \int |F|^2 \, d\mu_y - |m(y)|^2 = \int |F - m(y)|^2 \, d\mu_y.$$

Thus, for $y \in G$, the point y is predictable if and only if $F = m(y)$ holds μ_y-almost surely.

Suppose that (7) holds. By Definition 1.26(iii) applied to $A = X \setminus X_h$, we have $\mu_y(X \setminus X_h) = 0$ for ν-a.e. y, so for ν-a.e. $y \in G$ the measure μ_y is concentrated on $X_h \cap \phi^{-1}(y)$. By (7), F is constant on this set, so $F = m(y)$ μ_y-almost surely, and y is predictable.

Conversely, suppose that $\sigma(y) = 0$ for ν-a.e. y. By Definition 1.26(iii), the function $y \mapsto \mu_y(A)$ is ν-measurable for every Borel set A, hence (approximating F uniformly by simple functions) so is $y \mapsto m(y)$. Therefore there is a Borel map $\tilde{m} \colon \mathbb{R}^k \to \mathbb{R}^k$ with $\tilde{m} = m$ ν-almost everywhere. Let

$$X_h = \{x \in X : F(x) = \tilde{m}(\phi(x))\}.$$

This is a Borel set. For ν-a.e. y we have $y \in G$, $\tilde{m}(y) = m(y)$ and $\sigma(y) = 0$, and then $\mu_y(X_h) = \mu_y(\{x \in \phi^{-1}(y) : F(x) = m(y)\}) = 1$. By Definition 1.26(iii), $\mu(X_h) = \int \mu_y(X_h) \, d\nu(y) = 1$. Finally, if $x_1, x_2 \in X_h$ and $\phi(x_1) = \phi(x_2) = y$, then $F(x_1) = \tilde{m}(y) = F(x_2)$, which is (7). If ϕ is injective on a set of full measure, then (7) holds trivially on this set. □

Reference: [BGŚ22, Lemma 3.2, Cor. 3.3], [BGŚ24, Thm. 1.17].

Relation to Theorem 3.8. Embedding the manifold into $\mathbb{R}^N$ (Theorem 1.3), Theorem 3.8(ii) shows that SSOY Conjecture 1 holds with $\dim_{\mathrm{H}} \mu$ in place of $\mathrm{ID}(\mu)$, for arbitrary Borel probability measures (here generic observables are understood in the sense of prevalence in the space of Lipschitz observables). For ergodic measures, this gives the conjecture in its original form.

Corollary 3.15 (SSOY predictability conjecture for ergodic measures)**.** *Let $X \subset \mathbb{R}^N$ be compact, $T \colon X \to X$ Lipschitz and μ a T-invariant ergodic Borel probability measure on X. If $k > \underline{\mathrm{ID}}(\mu)$, then a prevalent Lipschitz observable $h \in \mathrm{Lip}(X, \mathbb{R})$ is almost surely k-predictable. In particular, SSOY Conjecture 1 holds for ergodic natural measures.*

Proof. For T-invariant ergodic measures $\dim_{\mathrm{H}} \mu \le \underline{\mathrm{ID}}(\mu)$ [BGŚ22, Prop. 2.1], so the assertion follows from Theorem 3.8(ii). □

Remark 3.16 (Counterexample [BGŚ22]). The ergodicity assumption in Corollary 3.15 cannot be dropped: there is a C^∞ diffeomorphism of $S^2 \times S^1$ with a natural measure μ with $\mathrm{ID}(\mu) < 1$ such that a prevalent Lipschitz observable h is not almost surely 1-predictable [BGŚ22, Thm. 1.11]. Hence SSOY Conjecture 1 fails in general. (By Theorem 3.8(ii), necessarily $\dim_{\mathrm{H}} \mu \ge 1 > \mathrm{ID}(\mu)$.)

Concerning SSOY Conjecture 2, Theorem 3.8(iv) gives (i) with $\dim_{\mathrm{H}} \mu$ in place of D (for ergodic μ), and Theorem 3.8(iii) gives the upper bound in (ii) with $\overline{\dim}_{\mathrm{B}} X$ in place of D. The latter cannot be replaced by $\dim_{\mathrm{H}} \mu$ or $\mathrm{ID}(\mu)$ for general measures [BGŚ24, Prop. 8.3]. Assertion (iii) holds with $\underline{\dim}_{\mathrm{B}} X$ in place of D: for a prevalent Lipschitz observable h, if $k > 2\underline{\dim}_{\mathrm{B}} X$ or $\phi_{h,k}$ is injective, then for every $\delta > 0$ the set $\{x : \sigma_\varepsilon(\phi_{h,k}(x)) > \delta\}$ is empty for small ε [BGŚ24, Thm. 1.7(ii)].

3.5 Predictability and the Farmer–Sidorowich algorithm

In practice, predictions are made from a single (long) time series. Let $x \in X$ and consider the delay vectors along its orbit,

$$z_i = z_i(x) = \phi(T^i x) = \big(h(T^i x), \dots, h(T^{i+k-1}x)\big), \qquad i = 0, 1, 2, \dots$$

The *Farmer–Sidorowich algorithm* [FS87] predicts the next delay vector by averaging the successors of those past delay vectors which are ε-close to the current one. More precisely, for $y \in \mathbb{R}^k$, $n \in \mathbb{N}$ and $\varepsilon > 0$ let

$$I_n = I_n(x, y, \varepsilon) = \{0 \le i < n : \ z_i \in B(y, \varepsilon)\}$$

and, if $I_n \neq \emptyset$,

$$\mathrm{Pred}_{x,n,\varepsilon}(y) = \frac{1}{\#I_n} \sum_{i \in I_n} z_{i+1}, \qquad \mathrm{Var}_{x,n,\varepsilon}(y) = \frac{1}{\#I_n} \sum_{i \in I_n} \big|z_{i+1} - \mathrm{Pred}_{x,n,\varepsilon}(y)\big|^2.$$

Knowing $z_0, \dots, z_n$, the algorithm predicts the next delay vector z_{n+1} as $\mathrm{Pred}_{x,n,\varepsilon}(z_n)$ (see Figure 10), and $\mathrm{Var}_{x,n,\varepsilon}(z_n)$ is the empirical variance of this prediction. The Farmer–Sidorowich algorithm and its variants, such as the simplex algorithm [SM90], are standard tools of nonparametric prediction. For ergodic measures, the quantities from Definition 3.6 are the limits of these empirical quantities as the length of the time series tends to infinity.

Proposition 3.17. *Let $X \subset \mathbb{R}^N$ be compact, $T\colon X \to X$ and $h\colon X \to \mathbb{R}$ continuous, μ a T-invariant ergodic Borel probability measure on X and $\phi = \phi_{h,k}$. Let $y \in \mathbb{R}^k$ and $\varepsilon > 0$ be such that $\mu(\phi^{-1}(B(y,\varepsilon))) > 0$. Then for μ-a.e. $x \in X$,*

$$\lim_{n\to\infty} \mathrm{Pred}_{x,n,\varepsilon}(y) = m_\varepsilon(y) \qquad \textit{and} \qquad \lim_{n\to\infty} \mathrm{Var}_{x,n,\varepsilon}(y) = \sigma_\varepsilon(y)^2.$$

Proof. Let $A = \phi^{-1}(B(y,\varepsilon))$ and $F = \phi \circ T$, a continuous, hence bounded map. Then $i \in I_n$ if and only if $T^i x \in A$, and $z_{i+1} = F(T^i x)$. Hence

$$\mathrm{Pred}_{x,n,\varepsilon}(y) = \frac{\frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_A(T^i x)\, F(T^i x)}{\frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_A(T^i x)}, \qquad \frac{1}{\#I_n}\sum_{i\in I_n} |z_{i+1}|^2 = \frac{\frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_A(T^i x)\, |F(T^i x)|^2}{\frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_A(T^i x)}.$$

By the Birkhoff ergodic theorem, applied to the bounded functions $\mathbf{1}_A$, $\mathbf{1}_A F$ (coordinatewise) and $\mathbf{1}_A |F|^2$, for μ-a.e. x the averages over $0 \le i < n$ converge to the corresponding integrals with respect to μ. As $\mu(A) > 0$, we obtain

$$\mathrm{Pred}_{x,n,\varepsilon}(y) \longrightarrow \frac{1}{\mu(A)} \int_A F \, d\mu = m_\varepsilon(y),$$
$$\frac{1}{\#I_n}\sum_{i\in I_n} |z_{i+1}|^2 \longrightarrow \frac{1}{\mu(A)} \int_A |F|^2 \, d\mu = \int |F|^2 \, d\mu_{y,\varepsilon}.$$

Expanding the square, $\mathrm{Var}_{x,n,\varepsilon}(y) = \frac{1}{\#I_n}\sum_{i\in I_n} |z_{i+1}|^2 - |\mathrm{Pred}_{x,n,\varepsilon}(y)|^2$, and the same identity holds for $\sigma_\varepsilon(y)^2$ and $\mu_{y,\varepsilon}$. Hence $\mathrm{Var}_{x,n,\varepsilon}(y) \to \sigma_\varepsilon(y)^2$. □

Reference: [BGŚ24, Prop. 1.4].

Thus $m_\varepsilon(y)$ is the long-time prediction of the Farmer–Sidorowich algorithm at y with accuracy ε, and $\sigma_\varepsilon(y)$ is its error. A point y is predictable if and only if this error vanishes as $\varepsilon \to 0$. Combining Proposition 3.17 with Theorem 3.8(ii) and Proposition 3.7, we obtain that for typical observables the algorithm is asymptotically exact, as soon as $k > \dim_{\mathrm{H}} \mu$.

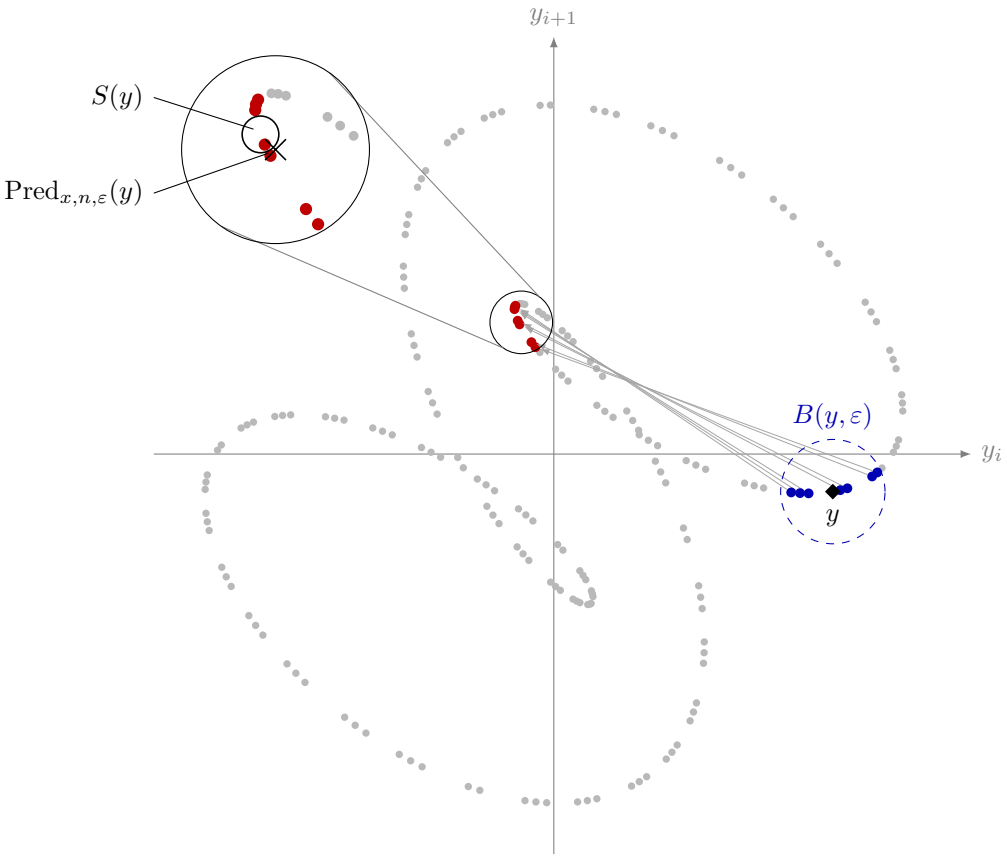


Figure 10: The Farmer–Sidorowich algorithm for the time series from Figure 7 (rotation by 43°, $k = 2$, $n = 193$, $\varepsilon = 0.3$). The grey points are the delay vectors $z_i = (y_i, y_{i+1})$, $i < n$, of a single orbit. To predict z_{n+1}, one finds the past delay vectors z_i, $i \in I_n$, which lie in the ball $B(y, \varepsilon)$ around the current delay vector $y = z_n$ (blue), and averages their successors z_{i+1} (red). The average $\mathrm{Pred}_{x,n,\varepsilon}(y)$ (cross) is close to the true next delay vector $z_{n+1} = \phi(T^{n+1}x) = S(y)$ (circle), where S is the prediction map from Proposition 3.2 (here ϕ is injective near the orbit points involved); see the inset, which shows the small circle around the prediction magnified three times.

Corollary 3.18. *Let $X \subset \mathbb{R}^N$ be compact, $T\colon X \to X$ Lipschitz, μ a T-invariant ergodic Borel probability measure on X, $k > \dim_{\mathrm{H}} \mu$ and $\varepsilon_j \searrow 0$. Then for a prevalent Lipschitz observable $h\colon X \to \mathbb{R}$, with $\phi = \phi_{h,k}$, the following holds. By Theorem 3.8(ii) and Proposition 3.7, there are a Borel set $X_h \subset X$ of full μ-measure and a prediction map S with*

$$\phi \circ T = S \circ \phi \qquad \text{on } X_h.$$

Then for μ-a.e. $x \in X$ and $\phi\mu$-a.e. $y \in \mathbb{R}^k$,

$$\lim_{j\to\infty} \lim_{n\to\infty} \mathrm{Pred}_{x,n,\varepsilon_j}(y) = S(y) \qquad \text{and} \qquad \lim_{j\to\infty} \lim_{n\to\infty} \mathrm{Var}_{x,n,\varepsilon_j}(y) = 0.$$

In other words, the Farmer–Sidorowich algorithm recovers the prediction map S given by the almost sure predictability of h.

Sketch of proof. As in the proof of Proposition 3.7, for $\phi\mu$-a.e. y the conditional measure μ_y is concentrated on $X_h \cap \phi^{-1}(y)$, where $\phi \circ T = S(y)$, so $m_\varepsilon(y) \to S(y)$ and $\sigma_\varepsilon(y) \to 0$ as $\varepsilon \to 0$. By Proposition 3.17 and Fubini's theorem (applied to $\mu \otimes \phi\mu$, for each of the countably many ε_j), for μ-a.e. x and $\phi\mu$-a.e. y we have $\mathrm{Pred}_{x,n,\varepsilon_j}(y) \to m_{\varepsilon_j}(y)$ and $\mathrm{Var}_{x,n,\varepsilon_j}(y) \to \sigma_{\varepsilon_j}(y)^2$ as $n \to \infty$, for every j. □

Reference: [BGŚ24, Cor. 1.18].

On the other hand, if $k < \dim_{\mathrm{H}} \mu$, then by Theorem 3.8(iv) the error of the Farmer–Sidorowich algorithm does not vanish on a set of positive measure, for a prevalent observable.

Part 4: Applications to fractal geometry

4.1 Iterated function systems

Many fractal sets are *self-similar*: they consist of finitely many smaller copies of themselves. Such sets, and natural measures on them, are generated by iterated function systems. Every point of the fractal is coded by an infinite sequence of symbols, and the coding map from the symbolic space onto the fractal is the main object of this part. It turns out that, applying the methods of embedding theory to this map, one can obtain interesting results on the geometry of fractal measures.

Definition 4.1 (Iterated function system)**.** An *iterated function system* (IFS) on a closed set $V \subset \mathbb{R}^d$ is a finite family $\mathcal{F} = \{f_i\}_{i\in\mathcal{A}}$, $\#\mathcal{A} \geq 2$, of contractions $f_i \colon V \to V$, i.e. there is $r < 1$ with $|f_i(x) - f_i(y)| \leq r|x-y|$ for all $x, y \in V$ and $i \in \mathcal{A}$. By Hutchinson's theorem [Hut81], there is a unique non-empty compact set $\Lambda = \Lambda_{\mathcal{F}} \subset V$ such that

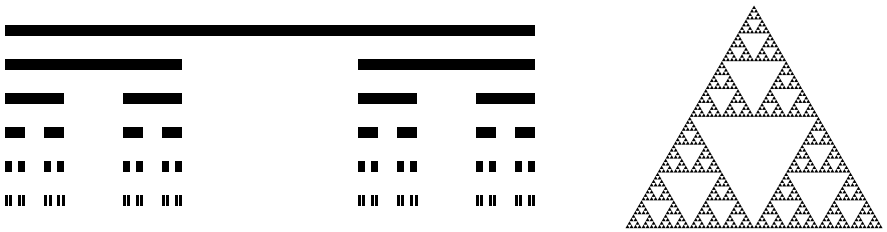
$$\Lambda = \bigcup_{i\in\mathcal{A}} f_i(\Lambda);$$

it is called the *attractor* of $\mathcal{F}$.

The most important classes are *self-similar* IFS, consisting of similarities $f_i(x) = r_i O_i x + t_i$ with $r_i \in (0,1)$, $O_i \in O(d)$ and $t_i \in \mathbb{R}^d$; *self-conformal* IFS, consisting of $C^{1+\theta}$ conformal contractions defined on a neighbourhood of V; and *self-affine* IFS, consisting of affine contractions. In this part we state all results for self-similar IFS. Everything we do below can be extended to self-conformal IFS as well (using the bounded distortion property), see e.g. [BSS23, BSŚ24].

Example 4.2. (i) The middle-third Cantor set is the attractor of $\{x \mapsto x/3,\ x \mapsto x/3 + 2/3\}$ on $\mathbb{R}$.

(ii) The Sierpiński gasket is the attractor of $\{x \mapsto x/2 + t_i\}_{i=1,2,3}$ on $\mathbb{R}^2$, where $t_1 = (0,0)$, $t_2 = (\frac{1}{2}, 0)$ and $t_3 = (\frac{1}{4}, \frac{\sqrt{3}}{4})$.
See Figure 11.



(a) (b)

Figure 11: Stages of the construction of the fractals from Example 4.2: (a) the middle-third Cantor set, (b) the Sierpiński gasket.

Symbolic coding. Let $\Sigma = \mathcal{A}^{\mathbb{N}}$ be the space of infinite words $\omega = \omega_1\omega_2\ldots$ over the alphabet $\mathcal{A}$, with the left shift $\sigma(\omega_1\omega_2\ldots) = \omega_2\omega_3\ldots$. For a finite word $u = u_1\ldots u_n \in \mathcal{A}^n$ we write $|u| = n$, $f_u = f_{u_1}\circ\cdots\circ f_{u_n}$ (with $f_\emptyset = \mathrm{id}$), and $[u] = \{\omega\in\Sigma : \omega_1\ldots\omega_n = u\}$ for the corresponding *cylinder*. For $\omega\in\Sigma$ let $\omega|_n = \omega_1\ldots\omega_n$, and for $\omega\neq\tau$ let $\omega\wedge\tau$ be their longest common prefix.

Lemma 4.3 (Natural projection)**.** *Let $\mathcal{F}$ be an IFS on V and $x_0\in V$. For every $\omega\in\Sigma$ the limit*

$$\pi(\omega) = \pi_{\mathcal{F}}(\omega) = \lim_{n\to\infty} f_{\omega|_n}(x_0)$$

exists and does not depend on x_0. The natural projection $\pi\colon\Sigma\to\mathbb{R}^d$ *is continuous, $\pi(\omega) = f_{\omega|_n}(\pi(\sigma^n\omega))$ for every n, $\pi([u]) = f_u(\Lambda)$ for every finite word u, and $\pi(\Sigma) = \Lambda$.*

Reference: [Hut81]; see e.g. [BSS23].

Thus every point of Λ has at least one *address* $\omega\in\Sigma$, and π is injective if and only if every point has exactly one address.

Measures. For a probability vector $p = (p_i)_{i\in\mathcal{A}}$, the Bernoulli measure $p^{\mathbb{N}}$ on Σ is shift-invariant and ergodic. Its image $\nu_p = \pi p^{\mathbb{N}}$ is the unique Borel probability measure satisfying $\nu_p = \sum_i p_i\, f_i\nu_p$; it is called a *self-similar* measure. More generally, we consider the set $\mathcal{E}(\Sigma)$ of shift-invariant ergodic Borel probability measures on Σ and their images $\pi\mu$, $\mu\in\mathcal{E}(\Sigma)$, which are ergodic measures on Λ.

Example 4.4 (Bernoulli convolutions)**.** Let $\mathcal{F}_\lambda = \{x\mapsto\lambda x-1,\ x\mapsto\lambda x+1\}$ and $p = (\frac12,\frac12)$. The self-similar measure $\nu_\lambda = \pi p^{\mathbb{N}}$ is the distribution of the random series $\sum_{n\ge0}\pm\lambda^n$ with independent fair signs. For $\lambda\in[\frac12,1)$ the attractor is the interval $[-\frac{1}{1-\lambda},\frac{1}{1-\lambda}]$, but ν_λ can be singular: Erdős [Erd39] showed that this is the case if $1/\lambda$ is a Pisot number. On the other hand, ν_λ is absolutely continuous for all $\lambda\in(\frac12,1)$ outside a set of zero Hausdorff dimension [Shm14]. For the dimension, $\dim_{\mathrm{H}}\nu_\lambda = 1$ for all transcendental $\lambda\in(\frac12,1)$ [Var19], see also [Hoc14]. See Figure 12.

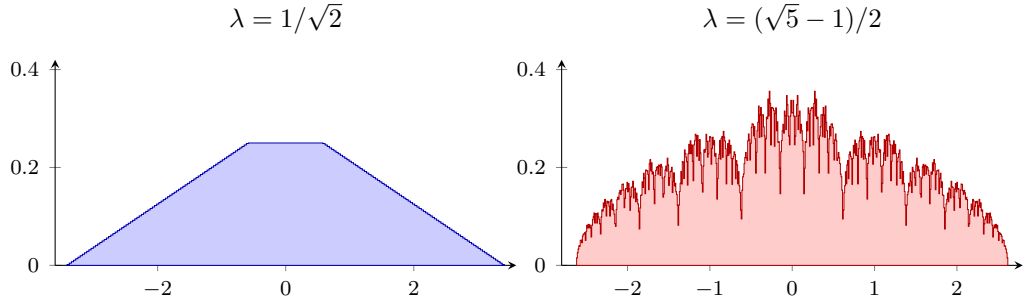


Figure 12: Histograms of Bernoulli convolutions ν_λ, computed from the 2^{22} sums $\sum_{n<22}\pm\lambda^n$. Left: for $\lambda = 1/\sqrt{2}$, ν_λ is absolutely continuous. Right: for the reciprocal $\lambda = (\sqrt{5}-1)/2$ of the golden ratio (a Pisot number), ν_λ is singular [Erd39].

Separation condition. The behaviour of π depends on the overlaps between the pieces $f_i(\Lambda) = \pi([i])$. An IFS satisfies the *strong separation condition* (SSC) if the sets $f_i(\Lambda)$, $i\in\mathcal{A}$, are pairwise disjoint. The middle-third Cantor set satisfies the SSC, the pieces of the Sierpiński gasket touch, and Bernoulli convolutions with $\lambda > \frac12$ have overlaps.

Symbolic metric, entropy and Lyapunov exponent. The natural metric on Σ is adapted to the contraction rates of the IFS.

Definition 4.5 (Symbolic metric, entropy, Lyapunov exponent)**.** Let $\mathcal{F} = \{f_i(x) = r_i O_i x + t_i\}_{i\in\mathcal{A}}$ be a self-similar IFS. For a finite word u write $r_u = r_{u_1}\cdots r_{u_n}$ (with $r_\emptyset = 1$), and let

$$\rho_{\mathcal{F}}(\omega,\tau) = \begin{cases} r_{\omega\wedge\tau}, & \omega \neq \tau, \\ 0, & \omega = \tau. \end{cases}$$

This is a metric on Σ (in fact an ultrametric) which generates the product topology (a cylinder $[u]$ has $\rho_{\mathcal{F}}$-diameter r_u). For $\mu \in \mathcal{E}(\Sigma)$ let

$$h(\mu) = \lim_{n\to\infty} -\frac{1}{n}\sum_{|u|=n} \mu[u]\log\mu[u], \qquad \chi_{\mathcal{F}}(\mu) = -\sum_{i\in\mathcal{A}} \mu[i]\log r_i$$

be the *entropy* and the *Lyapunov exponent* of μ. The *similarity dimension* $s(\mathcal{F})$ is the unique solution s of $\sum_{i\in\mathcal{A}} r_i^s = 1$.

For instance, $s(\mathcal{F}) = \log 2/\log 3$ for the middle-third Cantor set and $s(\mathcal{F}) = \log 3/\log 2$ for the Sierpiński gasket.

Recall that by the Shannon–McMillan–Breiman theorem, for $\mu \in \mathcal{E}(\Sigma)$,

$$-\frac{1}{n}\log\mu[\omega|_n] \longrightarrow h(\mu) \qquad \text{for } \mu\text{-a.e. } \omega \in \Sigma, \tag{10}$$

see e.g. [BSS23]. Dimensions of sets and measures in the metric space $(\Sigma, \rho_{\mathcal{F}})$ are defined as in Section 1.3.

Proposition 4.6. *Let $\mathcal{F} = \{f_i(x) = r_i O_i x + t_i\}_{i\in\mathcal{A}}$ be a self-similar IFS on $\mathbb{R}^d$, $\rho = \rho_{\mathcal{F}}$ and $s = s(\mathcal{F})$.*

(i) $\pi\colon (\Sigma,\rho) \to \mathbb{R}^d$ is Lipschitz. It is injective if and only if the SSC holds, and in this case it is bi-Lipschitz.

(ii) Every $\mu \in \mathcal{E}(\Sigma)$ is exact dimensional in (Σ,ρ) with

$$\dim(\mu,\rho) = \frac{h(\mu)}{\chi_{\mathcal{F}}(\mu)}.$$

(iii) $\dim_{\mathrm{H}}(\Sigma,\rho) = \dim_{\mathrm{A}}(\Sigma,\rho) = s$.

Consequently, $\dim_{\mathrm{H}}\pi\mu \le \min\{d, h(\mu)/\chi_{\mathcal{F}}(\mu)\}$ for $\mu \in \mathcal{E}(\Sigma)$ and $\dim_{\mathrm{H}}\Lambda \le \min\{d,s\}$, with equalities under the SSC.

Proof. (i) Let $\omega \neq \tau$, $u = \omega\wedge\tau$ and $n = |u|$. By Lemma 4.3, $\pi(\omega) = f_u(a)$ and $\pi(\tau) = f_u(b)$ with $a = \pi(\sigma^n\omega)$, $b = \pi(\sigma^n\tau) \in \Lambda$, so

$$|\pi(\omega) - \pi(\tau)| = r_u|a-b| \le \operatorname{diam}(\Lambda)\,\rho(\omega,\tau).$$

Here $a \in f_i(\Lambda)$ and $b \in f_j(\Lambda)$ with $i = \omega_{n+1} \neq j = \tau_{n+1}$. If the SSC holds, then $\delta = \min_{i\neq j}\operatorname{dist}(f_i(\Lambda), f_j(\Lambda)) > 0$, so $|\pi(\omega)-\pi(\tau)| \ge \delta r_u = \delta\rho(\omega,\tau)$, and π is bi-Lipschitz. Conversely, if $z \in f_i(\Lambda)\cap f_j(\Lambda)$ for some $i \neq j$, then, as $f_i(\Lambda) = \pi([i])$, we have $z = \pi(\omega) = \pi(\tau)$ for some $\omega \in [i]$, $\tau \in [j]$, so π is not injective.

(ii) Let $r_{\min} = \min_i r_i$. For $\omega \in \Sigma$ and $R \in (0,1)$ let $n = n(\omega,R)$ be the smallest n with $r_{\omega|_n} \le R$. Since $\rho(\omega,\tau) = r_{\omega\wedge\tau}$ and $r_{\omega|_m}$ is decreasing in m, the closed ball $\overline{B}(\omega,R)$ in (Σ,ρ)

is the cylinder $[\omega|_n]$. Moreover, $r_{\omega|_n} \le R < r_{\omega|_{n-1}} \le r_{\omega|_n}/r_{\min}$, so $\log R = \log r_{\omega|_n} + O(1)$, and $n \to \infty$ as $R \to 0$. Hence

$$\frac{\log \mu(\overline{B}(\omega, R))}{\log R} = \frac{-\frac{1}{n} \log \mu[\omega|_n]}{-\frac{1}{n} \log r_{\omega|_n} + O(\frac{1}{n})}.$$

By (10), the numerator tends to $h(\mu)$ for μ-a.e. ω. The denominator tends to $\chi_{\mathcal{F}}(\mu)$, since $-\frac{1}{n} \log r_{\omega|_n} = \frac{1}{n} \sum_{j=0}^{n-1} g(\sigma^j \omega)$ for $g(\omega) = -\log r_{\omega_1}$, and $\frac{1}{n} \sum_{j=0}^{n-1} g(\sigma^j \omega) \to \int g \, d\mu = \chi_{\mathcal{F}}(\mu)$ for μ-a.e. ω by the Birkhoff ergodic theorem. As $\chi_{\mathcal{F}}(\mu) > 0$, we obtain $\underline{d}(\mu, \omega) = \overline{d}(\mu, \omega) = h(\mu)/\chi_{\mathcal{F}}(\mu)$ for μ-a.e. ω.

(iii) *Upper bound.* Let $\omega \in \Sigma$ and $0 < r < R < 1$. By (ii), $\overline{B}(\omega, R) = [v]$ for some word v with $r_v \le R$. If $r \ge r_v$, then $[v]$ has diameter at most r. Otherwise, let W be the (finite) set of words w beginning with v such that $r_w \le r < r_{w^-}$, where w^- is w with the last letter removed. The cylinders $[w]$, $w \in W$, partition $[v]$ and have diameter at most r. As $\sum_i r_i^s = 1$, replacing a word u by the words ui, $i \in \mathcal{A}$, does not change the sum of r_u^s, so $\sum_{w \in W} r_w^s = r_v^s$. Since $r_w > r_{\min} r$ for $w \in W$, we get $\#W \, (r_{\min} r)^s \le r_v^s \le R^s$. Hence $\overline{B}(\omega, R)$ can be covered by at most $r_{\min}^{-s} (R/r)^s$ sets of diameter at most r, so $\dim_{\mathrm{A}}(\Sigma, \rho) \le s$.

Lower bound. Let $\mu_s = p^{\mathbb{N}}$ with $p_i = r_i^s$. Then $h(\mu_s) = -\sum_i p_i \log p_i = -s \sum_i r_i^s \log r_i = s \, \chi_{\mathcal{F}}(\mu_s)$, so by (ii), $\underline{d}(\mu_s, \omega) = s$ for μ_s-a.e. ω. By Proposition 1.11 (which holds in every metric space), $\dim_{\mathrm{H}}(\Sigma, \rho) \ge \dim_{\mathrm{H}} \mu_s = s$. As $\dim_{\mathrm{H}} \le \dim_{\mathrm{A}}$, we obtain (iii).

Consequences. The upper bounds hold since π is Lipschitz by (i), so it cannot increase the Hausdorff dimension (the bounds by d are trivial). Under the SSC, π is bi-Lipschitz by (i), so the dimensions are preserved. □

Reference: [Hut81], [Fal14, Ch. 9], [BSS23, Ch. 14].

Thus, if π is bi-Lipschitz, the dimensions of Λ and of $\pi\mu$ are given by the symbolic quantities $s(\mathcal{F})$ and $h(\mu)/\chi_{\mathcal{F}}(\mu)$. Without separation the dimension may drop, and the question of when the dimension is preserved is a question about embeddings.

Natural projections as embeddings. The natural projection $\pi \colon (\Sigma, \rho_{\mathcal{F}}) \to \mathbb{R}^d$ is a Lipschitz map from a metric space of dimension $s(\mathcal{F})$, and $\mu \in \mathcal{E}(\Sigma)$ is a measure of dimension $h(\mu)/\chi_{\mathcal{F}}(\mu)$. The questions of Parts 1 and 2 become: Is π injective / μ-a.s. injective? How regular is the inverse? The following theorem shows that for self-similar measures, almost sure injectivity of π is exactly the condition for the dimension to be preserved. As we have seen in Example 1.29, this is not the case for general maps: an injective (even Lipschitz) map can decrease the dimension. Here, the self-similar structure of π and μ is crucial.

Theorem 4.7 (Feng–Hu)**.** *Let $\mathcal{F}$ be a self-similar IFS and $\mu \in \mathcal{E}(\Sigma)$. Then $\pi\mu$ is exact dimensional. Moreover, π is μ-a.s. injective if and only if*

$$\dim \pi\mu = \frac{h(\mu)}{\chi_{\mathcal{F}}(\mu)}.$$

Reference: [FH09], [KS19, Appendix], [Fen23, Cor. 4.7]; see [BSS23, Thms. 11.1.2, 11.3.1, Prop. 11.3.6] and [BGŚ26b, Section 1.6].

4.2 Parametrized families of IFS

For a given IFS with overlaps, studying the geometry of attractors and invariant measures is hard. As in Parts 1–3, the transversality method gives results for typical members of parametrized families of IFS. Projections of self-similar sets form a basic example.

Example 4.8 (Orthogonal projections of self-similar sets)**.** Let $\mathcal{F} = \{x \mapsto rx + t_i\}$ on $\mathbb{R}^n$ and $V \in \mathrm{Gr}(k, n)$. Then $P_V \circ \pi_{\mathcal{F}} = \pi_{\mathcal{F}_V}$ for the IFS $\mathcal{F}_V = \{y \mapsto ry + P_V t_i\}$ on V. Thus projections of self-similar sets are self-similar sets, typically with overlaps, and projection theorems for $P_V \circ \pi_{\mathcal{F}}$ are statements about typical members of the family $\{\mathcal{F}_V\}$ (cf. [Ken97]).

Transversality. Let $\{\mathcal{F}^\lambda\}_{\lambda\in U}$ be a family of self-similar IFS $\mathcal{F}^\lambda = \{f_i^r(x) = r_i(\lambda)O_i(\lambda)x + t_i(\lambda)\}_{i\in\mathcal{A}}$ on $\mathbb{R}^d$ depending continuously on a parameter λ in a metric space U, with $0 < r_{\min} \le r_i(\lambda) \le r_{\max} < 1$, and let η be a finite measure on U. Write $\pi_\lambda = \pi_{\mathcal{F}^\lambda}$, $\rho_\lambda = \rho_{\mathcal{F}^\lambda}$, χ_λ.

Definition 4.9 (Transversality condition)**.** The family satisfies the *transversality condition* if there is $C > 0$ with

$$\eta(\{\lambda \in U : |\pi_\lambda(\omega) - \pi_\lambda(\tau)| \le r\}) \le Cr^d \qquad \text{for all } r > 0 \text{ and } \omega, \tau \in \Sigma \text{ with } \omega_1 \ne \tau_1.$$

As $|\pi_\lambda(\omega) - \pi_\lambda(\tau)| = \rho_\lambda(\omega,\tau)\,|\pi_\lambda(\sigma^n\omega) - \pi_\lambda(\sigma^n\tau)|$ for $n = |\omega \wedge \tau|$, this implies $\eta(\{\lambda : |\pi_\lambda(\omega) - \pi_\lambda(\tau)| \le r\rho_\lambda(\omega,\tau)\}) \le Cr^d$ for all $\omega \ne \tau$.

The transversality condition is the analogue of the key estimates from the previous parts: Lemma 1.15 for orthogonal projections, Lemma 1.33 for linear maps and Lemmas 3.9, 3.11 for delay-coordinate maps (cf. Remark 1.16). Unlike for orthogonal projections, it does not come for free. For delay-coordinate maps we were able to obtain it after some additional work (and under assumptions on periodic points). Here the situation is even worse: transversality may fail on some range of parameters, see Example 4.10(iii).

Example 4.10 (Transversal families)**.** (i) Orthogonal projections $V \mapsto P_V \circ \pi_{\mathcal{F}}$ of a self-similar set with SSC (Remark 1.16).

(ii) Translation families $\{x \mapsto r_iO_ix + t_i\}_{i\in\mathcal{A}}$ with fixed r_i, O_i and $t = (t_i)$ in an open set of parameters, provided $\max_{i\ne j}(r_i + r_j) < 1$ [BSS23, Thm. 14.5.2], [BSŚ24, Ex. 3.8]. For $x \mapsto r_ix + t_i$ on $\mathbb{R}$ this holds whenever $s_0 = s(\mathcal{F}) < 1$.

(iii) Bernoulli convolutions: transversality holds on $(0, 0.649]$ [Sol95, PS96], [BSS23, Cor. 6.4.5]. The maximal transversality interval is approximately $(0, 0.6684]$ [SS06], and transversality fails on $(2^{-1/2}, 1)$ [Sol98]; see [BSS23, Rem. 6.4.6].

Theorem 4.11 (Universal projection theorem [BSŚ24])**.** *Let $\{\mathcal{F}^\lambda\}_{\lambda\in U}$ be a continuous family of self-similar IFS on $\mathbb{R}^d$ satisfying the transversality condition. Then for η-a.e. λ, simultaneously for all $\mu \in \mathcal{E}(\Sigma)$:*

(i) $\dim_{\mathrm{H}} \pi_\lambda\mu = \min\{d, h(\mu)/\chi_\lambda(\mu)\}$*;*

(ii) if $s(\mathcal{F}^\lambda) < d$, then π_λ is μ-nearly bi-Lipschitz; in particular π_λ is μ-a.s. injective.

Theorem 4.11 is a special case of a general result based on transversality, concerning families of Lipschitz maps $\pi_\lambda\colon (X, \rho_\lambda) \to \mathbb{R}^d$ satisfying a transversality condition [BSŚ24, Thm. 2.7]. For a fixed measure μ, (i) is classical: it is due to Simon, Solomyak and Urbański [SSU01a, SSU01b], see also [BSS23, Thm. 14.4.2], and it is the analogue of Theorem 1.21 and the projection theorem (Theorem 1.28). Part (ii) is the analogue of Theorem 2.2(B)(iii), as $s(\mathcal{F}^\lambda) = \dim_{\mathrm{A}}(\Sigma, \rho_\lambda)$. The novelty is *universality*: the exceptional set of parameters does not depend on the measure. That is, there is a set $U_0 \subset U$ with $\eta(U \setminus U_0) = 0$ such that for every $\lambda \in U_0$, (i) and (ii) hold for *all* $\mu \in \mathcal{E}(\Sigma)$. This does not follow from the single-measure results, as $\mathcal{E}(\Sigma)$ is uncountable (also see Remark 4.12(i)).

In the self-similar setting, the dimension part (i) was known before: Jordan and Rapaport [JR21, Thm. 1.1] proved that if a self-similar IFS on $\mathbb{R}$ satisfies the exponential separation condition (which holds for typical parameters [Hoc14]), then $\dim_{\mathrm{H}} \pi_{\mathcal{F}}\mu = \min\{1, h(\mu)/\chi_{\mathcal{F}}(\mu)\}$ for all $\mu \in \mathcal{E}(\Sigma)$. This was not known in the general self-conformal case, to which Theorem 4.11 extends [BSŚ24]. The main application to multifractal analysis in Section 4.3 is possible because of part (ii), i.e. the Assouad dimension part.

Remark 4.12. (i) Universality is not automatic: for $V \in \mathrm{Gr}(1, 2)$ let μ_V be $\mathcal{H}^1$ on the unit segment perpendicular to V; then $\dim_{\mathrm{H}} \mu_V = 1$ but $\dim_{\mathrm{H}} P_V\mu_V = 0$ [BSŚ24, Ex. 3.2].

(ii) Applied to orthogonal projections: $\dim_{\mathrm{H}} P_V\mu = \min\{k, \dim_{\mathrm{H}} \mu\}$ for a.e. V simultaneously for all ergodic measures on a self-similar set with SSC, and for all Gibbs measures on an arbitrary self-similar set [BSŚ24, Thms. 3.5, 3.6] (cf. [HS12] for all projections under irrationality assumptions).

(iii) The general scheme also covers non-autonomous and random IFS [BSŚ24, Section 3.3].

4.3 Multifractal analysis

One can study much finer geometry of a measure than just its dimension. For instance, a self-similar measure ν is exact dimensional (Theorem 4.7), so its local dimension equals $\dim_{\mathrm{H}} \nu$ at ν-a.e. point. Other, non-typical scaling exponents α occur only on ν-null sets, which can nevertheless be large. The multifractal spectrum and the L^q-spectrum describe the distribution of these non-typical scaling exponents.

Definition 4.13 (Multifractal spectrum, L^q-spectrum)**.** For a finite Borel measure ν on $\mathbb{R}^d$ and $\alpha \geq 0$ let

$$E_\nu(\alpha) = \{x \in \operatorname{supp}\nu : \underline{d}(\nu, x) = \overline{d}(\nu, x) = \alpha\} \qquad \text{and} \qquad f_\nu(\alpha) = \dim_{\mathrm{H}} E_\nu(\alpha).$$

The function f_ν is the *multifractal spectrum* of ν. The *L^q-spectrum* of ν is

$$\tau_\nu(q) = \liminf_{r\to 0} \frac{\log \sup \sum_i \nu(B(x_i, r))^q}{\log r}, \qquad q \in \mathbb{R},$$

the supremum over families of disjoint balls of radius r centred in $\operatorname{supp}\nu$. The *multifractal formalism* holds at α if

$$f_\nu(\alpha) = \tau_\nu^*(\alpha) := \inf_{q\in\mathbb{R}} (\alpha q - \tau_\nu(q)),$$

so f_ν is the Legendre transform of τ_ν.

Self-similar measures. Let $\nu = \nu_{t,p}$ be the self-similar measure on $\mathbb{R}$ for $\mathcal{F}_t = \{x \mapsto r_i x + t_i\}_{i\in\mathcal{A}}$ and a probability vector p. For $q \in \mathbb{R}$ let $T(q)$ be the unique real number with

$$\sum_{i\in\mathcal{A}} p_i^q |r_i|^{-T(q)} = 1.$$

$T(q)$ is the expected value of $\tau_\nu(q)$, just as $h(\mu)/\chi_{\mathcal{F}}(\mu)$ is the expected value of $\dim_{\mathrm{H}} \pi_{\mathcal{F}}\mu$: in both cases equality holds under the SSC (Proposition 4.6 and Proposition 4.14 below). Moreover, T is concave and increasing, $T(1) = 0$ and $T(0) = -s_0$, where $s_0 = s(\mathcal{F}_t)$ is the similarity dimension, i.e. $\sum_{i\in\mathcal{A}} |r_i|^{s_0} = 1$. Let $T^*(\alpha) = \inf_{q\in\mathbb{R}}(\alpha q - T(q))$ be its Legendre transform. For

$$\alpha_{\min} = \min_{i\in\mathcal{A}} \frac{\log p_i}{\log |r_i|}, \qquad \alpha_{\max} = \max_{i\in\mathcal{A}} \frac{\log p_i}{\log |r_i|},$$

$T^*(\alpha)$ is finite if and only if $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, so this is the maximal range on which one can expect $f_\nu = T^*$ to hold. One always has $E_\nu(\alpha) = \emptyset$ for $\alpha > \alpha_{\max}$, and under the SSC also for $\alpha < \alpha_{\min}$.

Proposition 4.14 (Separated case)**.** *If $\mathcal{F}_t$ satisfies the SSC, then $\tau_\nu(q) = T(q)$ for all $q \in \mathbb{R}$ and $f_\nu(\alpha) = T^*(\alpha)$ for all $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ (so the multifractal formalism holds).*

Reference: [CM92], [AP96], [Ols95]; exposition [Fal97, Ch. 11], [BSS23, Ch. 5].

Much less is known if there are overlaps. Barral and Feng [BF21, Thm. 1.2, Rem. 7.3] proved that if $s_0 < 1$ and $\mathcal{F}_t$ satisfies the exponential separation condition (which holds for a.e. t), then $f_\nu(\alpha) = T^*(\alpha)$ on the increasing part of the spectrum, i.e. for

$$\alpha \in [\alpha_{\min}, \alpha_0], \qquad \text{where} \quad \alpha_0 = \frac{\sum_{i\in\mathcal{A}} |r_i|^{s_0} \log p_i}{\sum_{i\in\mathcal{A}} |r_i|^{s_0} \log |r_i|} = T'(0)$$

(corresponding to $q > 0$ in the definition of T^*). The decreasing part ($q < 0$) remained out of reach: L^q-spectra for negative q are notoriously difficult. Theorem 4.11(ii) allows to obtain the following result.

Theorem 4.15 (Full spectrum for typical translations [BSŚ24])**.** *Fix $r_i \in (-1, 1) \setminus \{0\}$, $i \in \mathcal{A}$, with $s_0 < 1$. Then for Lebesgue-a.e. $t \in \mathbb{R}^{\mathcal{A}}$,*

$$\dim_{\mathrm{H}} E_{\nu_{t,p}}(\alpha) = T^*(\alpha)$$

for every *probability vector p with $p \neq (|r_i|^{s_0})_i$ and* every *$\alpha \in [\alpha_{\min}, \alpha_{\max}]$.*

Sketch of proof. *Upper bound.* For $\alpha \in [\alpha_0, \alpha_{\max}]$ the bound

$$\dim_{\mathrm{H}} E_{\nu_{t,p}}(\alpha) \le T^*(\alpha)$$

holds for all t by a standard covering argument [BSŚ24, Lemma 8.2]. For $\alpha \in [\alpha_{\min}, \alpha_0]$ it follows from [BF21] for a.e. t.

Lower bound. Fix p and $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, let $q_\alpha \in \mathbb{R}$ satisfy $T'(q_\alpha) = \alpha$ and consider the Bernoulli measure

$$\mu_\alpha = \left(p_i^{q_\alpha} |r_i|^{-T(q_\alpha)}\right)_{i\in\mathcal{A}}^{\mathbb{N}}.$$

Let $\rho = \rho_{\mathcal{F}_t}$ (it does not depend on t). A direct computation gives

$$d(p^{\mathbb{N}}, \omega) = \alpha \ \text{ in } (\Sigma, \rho) \ \text{ for } \mu_\alpha\text{-a.e. } \omega \qquad \text{and} \qquad \frac{h(\mu_\alpha)}{\chi_{\mathcal{F}_t}(\mu_\alpha)} = \alpha q_\alpha - T(q_\alpha) = T^*(\alpha).$$

By Theorem 4.11 (transversality from Example 4.10(ii) and $s_0 = \dim_{\mathrm{A}}(\Sigma, \rho) < 1$), for a.e. t, *simultaneously for all α and p*, π_t is μ_α-nearly bi-Lipschitz and

$$\dim_{\mathrm{H}} \pi_t \mu_\alpha = T^*(\alpha).$$

As $\nu_{t,p} = \pi_t p^{\mathbb{N}}$, Lemma 2.8 gives

$$d(\nu_{t,p}, \pi_t \omega) = d(p^{\mathbb{N}}, \omega) = \alpha \qquad \text{for } \mu_\alpha\text{-a.e. } \omega,$$

hence $\pi_t \mu_\alpha(E_{\nu_{t,p}}(\alpha)) = 1$ and

$$\dim_{\mathrm{H}} E_{\nu_{t,p}}(\alpha) \ge \dim_{\mathrm{H}} \pi_t \mu_\alpha = T^*(\alpha).$$

See [BSŚ24, Thm. 1.1, Section 8]. □

Remark 4.16. (i) Universality is essential: the classical transversality results give a full-measure set of t depending on μ_α, and there are uncountably many α and p.

(ii) This is the first result on the multifractal spectrum of typical self-similar measures with overlaps covering the full range, including the decreasing part. It does not establish the full multifractal formalism: this requires $\tau_\nu(q) = T(q)$ also for $q < 0$.

(iii) The same methods in $\mathbb{R}^d$, $d \ge 2$ gives the multifractal spectrum for a.e. t, if $s_0 < d$ and $\max_{i\in\mathcal{A}} r_i < \frac{1}{2}$ [BSŚ24, Thm. 3.15]. The method extends to Birkhoff spectra of continuous potentials.

References

[Aba96] H. D. I. Abarbanel. *Analysis of Observed Chaotic Data*. Springer, New York, 1996.

[ABD+19] G. Alberti, H. Bölcskei, C. De Lellis, G. Koliander, and E. Riegler. Lossless analog compression. *IEEE Trans. Inform. Theory*, 65(11):7480–7513, 2019.

[AP96] M. Arbeiter and N. Patzschke. Random self-similar multifractals. *Math. Nachr.*, 181:5–42, 1996.

[Ass83] P. Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bull. Soc. Math. France*, 111:429–448, 1983.

[BAEFN93] A. Ben-Artzi, A. Eden, C. Foias, and B. Nicolaenko. Hölder continuity for the inverse of Mañé's projection. *J. Math. Anal. Appl.*, 178(1):22–29, 1993.

[BF21] J. Barral and D.-J. Feng. On multifractal formalism for self-similar measures with overlaps. *Math. Z.*, 298:359–383, 2021.

[BGŚ20] K. Barański, Y. Gutman, and A. Śpiewak. A probabilistic Takens theorem. *Nonlinearity*, 33(9):4940–4966, 2020.

[BGŚ22] K. Barański, Y. Gutman, and A. Śpiewak. On the Shroer–Sauer–Ott–Yorke predictability conjecture for time-delay embeddings. *Comm. Math. Phys.*, 391(2):609–641, 2022.

[BGŚ24] K. Barański, Y. Gutman, and A. Śpiewak. Prediction of dynamical systems from time-delayed measurements with self-intersections. *J. Math. Pures Appl.*, 186:103–149, 2024.

[BGŚ26a] K. Barański, Y. Gutman, and A. Śpiewak. Predicting dynamical systems with too few time-delay measurements: error estimates. *Trans. Amer. Math. Soc.*, 2026. To appear; arXiv:2401.15712.

[BGŚ26b] K. Barański, Y. Gutman, and A. Śpiewak. Regularity of almost-surely injective projections in Euclidean spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, 27(2):1109–1144, 2026.

[BSS23] B. Bárány, K. Simon, and B. Solomyak. *Self-similar and Self-affine Sets and Measures*, volume 276 of *Mathematical Surveys and Monographs*. Amer. Math. Soc., Providence, RI, 2023.

[BSŚ24] B. Bárány, K. Simon, and A. Śpiewak. Universal projection theorems with applications to multifractal analysis and the dimension of every ergodic measure on self-conformal sets simultaneously, 2024. Preprint, arXiv:2412.03529.

[Cab00] V. Caballero. On an embedding theorem. *Acta Math. Hungar.*, 88(4):269–278, 2000.

[Chr72] J. P. R. Christensen. On sets of Haar measure zero in abelian Polish groups. *Israel J. Math.*, 13:255–260, 1972.

[CM92] R. Cawley and R. D. Mauldin. Multifractal decompositions of Moran fractals. *Adv. Math.*, 92(2):196–236, 1992.

[CRT06] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[Don06] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[DPT25] G. C. David, P. Pankka, and J. T. Tyson. The Heinonen–Semmes problems after thirty years, 2025. Preprint, arXiv:2510.19179.

[ER85] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Modern Phys.*, 57(3):617–656, 1985.

[Erd39] P. Erdős. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.*, 61:974–976, 1939.

[Fal97] K. Falconer. *Techniques in Fractal Geometry*. John Wiley & Sons, Chichester, 1997.

[Fal14] K. Falconer. *Fractal Geometry. Mathematical Foundations and Applications*. John Wiley & Sons, Chichester, third edition, 2014.

[Fen23] D.-J. Feng. Dimension of invariant measures for affine iterated function systems. *Duke Math. J.*, 172(4):701–774, 2023.

[FFJ15] K. Falconer, J. Fraser, and X. Jin. Sixty years of fractal projections. In *Fractal Geometry and Stochastics V*, volume 70 of *Progr. Probab.*, pages 3–25. Birkhäuser/Springer, 2015.

[FH09] D.-J. Feng and H. Hu. Dimension theory of iterated function systems. *Comm. Pure Appl. Math.*, 62(11):1435–1500, 2009.

[FK20] J. M. Fraser and A. Käenmäki. Attainable values for the Assouad dimension of projections. *Proc. Amer. Math. Soc.*, 148:3393–3405, 2020.

[FR13] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[Fra20] J. M. Fraser. *Assouad Dimension and Fractal Geometry*, volume 222 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2020.

[FS87] J. D. Farmer and J. J. Sidorowich. Predicting chaotic time series. *Phys. Rev. Lett.*, 59:845–848, 1987.

[GG73] M. Golubitsky and V. Guillemin. *Stable Mappings and Their Singularities*, volume 14 of *Graduate Texts in Mathematics*. Springer-Verlag, New York–Heidelberg, 1973.

[GK19] B. C. Geiger and T. Koch. On the information dimension of stochastic processes. *IEEE Trans. Inform. Theory*, 65(10):6496–6518, 2019.

[GQS18] Y. Gutman, Y. Qiao, and G. Szabó. The embedding problem in topological dynamics and Takens' theorem. *Nonlinearity*, 31(2):597–620, 2018.

[GŚ19] Y. Gutman and A. Śpiewak. New uniform bounds for almost lossless analog compression. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1702–1706, 2019.

[GŚ20] Y. Gutman and A. Śpiewak. Metric mean dimension and analog compression. *IEEE Trans. Inform. Theory*, 66(11):6977–6998, 2020.

[GŚ26] Y. Gutman and A. Śpiewak. Optimal compressed sensing for mixing stochastic processes. *IEEE Trans. Inform. Theory*, 2026. To appear; arXiv:2507.23175.

[GT20] Y. Gutman and M. Tsukamoto. Embedding minimal dynamical systems into Hilbert cubes. *Invent. Math.*, 221(1):113–166, 2020.

[Gut16] Y. Gutman. Takens embedding theorem with a continuous observable. In *Ergodic Theory: Advances in Dynamical Systems*, pages 134–141. De Gruyter, Berlin, 2016.

[Hei01] J. Heinonen. *Lectures on Analysis on Metric Spaces*. Universitext. Springer, New York, 2001.

[Hei03] J. Heinonen. Geometric embeddings of metric spaces. Report 90, University of Jyväskylä, Department of Mathematics and Statistics, 2003.

[HGLS05] C.-H. Hsieh, S. M. Glaser, A. J. Lucas, and G. Sugihara. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature*, 435:336–340, 2005.

[Hir76] M. W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Springer, New York, 1976.

[HK99] B. R. Hunt and V. Yu. Kaloshin. Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces. *Nonlinearity*, 12(5):1263–1275, 1999.

[Hoc14] M. Hochman. On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)*, 180(2):773–822, 2014.

[HS12] M. Hochman and P. Shmerkin. Local entropy averages and projections of fractal measures. *Ann. of Math. (2)*, 175(3):1001–1059, 2012.

[HSY92] B. R. Hunt, T. Sauer, and J. A. Yorke. Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces. *Bull. Amer. Math. Soc. (N.S.)*, 27(2):217–238, 1992.

[HT94] X. Hu and S. J. Taylor. Fractal properties of products and projections of measures in $\mathbb{R}^d$. *Math. Proc. Cambridge Philos. Soc.*, 115(3):527–544, 1994.

[Huk06] J. P. Huke. Embedding nonlinear dynamical systems: a guide to Takens' theorem. MIMS EPrint 2006.26, University of Manchester, 2006.

[Hut81] J. E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.

[HW41] W. Hurewicz and H. Wallman. *Dimension Theory*, volume 4 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1941.

[JM98] M. Järvenpää and P. Mattila. Hausdorff and packing dimensions and sections of measures. *Mathematika*, 45(1):55–77, 1998.

[JMB14] S. Jalali, A. Maleki, and R. G. Baraniuk. Minimum complexity pursuit for universal compressed sensing. *IEEE Trans. Inform. Theory*, 60(4):2253–2268, 2014.

[JP17] S. Jalali and H. V. Poor. Universal compressed sensing for almost lossless recovery. *IEEE Trans. Inform. Theory*, 63(5):2933–2953, 2017.

[JR21] T. Jordan and A. Rapaport. Dimension of ergodic measures projected onto self-similar sets with overlaps. *Proc. Lond. Math. Soc. (3)*, 122(2):191–206, 2021.

[Kau68] R. Kaufman. On Hausdorff dimension of projections. *Mathematika*, 15:153–155, 1968.

[Ken97] R. Kenyon. Projecting the one-dimensional Sierpinski gasket. *Israel J. Math.*, 97:221–238, 1997.

[KS04] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, second edition, 2004.

[KS19] I. Kolossváry and K. Simon. Triangular Gatzouras–Lalley-type planar carpets with overlaps. *Nonlinearity*, 32(9):3294–3341, 2019.

[Laa02] T. J. Laakso. Plane with A_∞-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002.

[Lin99] E. Lindenstrauss. Mean dimension, small entropy factors and an embedding theorem. *Publ. Math. Inst. Hautes Études Sci.*, 89:227–262, 1999.

[LP01] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87:285–307, 2001.

[LT14] E. Lindenstrauss and M. Tsukamoto. Mean dimension and an embedding problem: an example. *Israel J. Math.*, 199(2):573–584, 2014.

[Mañ81] R. Mañé. On the dimension of the compact invariant sets of certain nonlinear maps. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Math.*, pages 230–242. Springer, Berlin, 1981.

[Mar54] J. M. Marstrand. Some fundamental geometrical properties of plane sets of fractional dimensions. *Proc. London Math. Soc. (3)*, 4:257–302, 1954.

[Mat75] P. Mattila. Hausdorff dimension, orthogonal projections and intersections with planes. *Ann. Acad. Sci. Fenn. Ser. A I Math.*, 1(2):227–244, 1975.

[Mat95] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[MGNS18] S. B. Munch, A. Giron-Nava, and G. Sugihara. Nonlinear dynamics and noise in fisheries recruitment: a global meta-analysis. *Fish and Fisheries*, 19(6):964–973, 2018.

[MS74] J. W. Milnor and J. D. Stasheff. *Characteristic Classes*, volume 76 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1974.

[NN12] A. Naor and O. Neiman. Assouad's theorem with dimension independent of the snowflaking. *Rev. Mat. Iberoam.*, 28(4):1123–1142, 2012.

[Ols95] L. Olsen. A multifractal formalism. *Adv. Math.*, 116(1):82–196, 1995.

[Ols02] E. J. Olson. Bouligand dimension and almost Lipschitz embeddings. *Pacific J. Math.*, 202(2):459–474, 2002.

[OY05] W. Ott and J. A. Yorke. Prevalence. *Bull. Amer. Math. Soc. (N.S.)*, 42(3):263–290, 2005.

[PCFS80] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45:712–716, 1980.

[PdM82] J. Palis, Jr. and W. de Melo. *Geometric Theory of Dynamical Systems. An Introduction*. Springer-Verlag, New York–Berlin, 1982.

[PS96] Y. Peres and B. Solomyak. Absolute continuity of Bernoulli convolutions, a simple proof. *Math. Res. Lett.*, 3(2):231–239, 1996.

[RJEP17] F. E. Rezagah, S. Jalali, E. Erkip, and H. V. Poor. Compression-based compressed sensing. *IEEE Trans. Inform. Theory*, 63(10):6735–6752, 2017.

[Rob05] J. C. Robinson. A topological delay embedding theorem for infinite-dimensional dynamical systems. *Nonlinearity*, 18(5):2135–2143, 2005.

[Rob11] J. C. Robinson. *Dimensions, Embeddings, and Attractors*, volume 186 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2011.

[Roh52] V. A. Rohlin. On the fundamental ideas of measure theory. *Amer. Math. Soc. Translation*, 1952(71), 1952.

[Sau06] T. D. Sauer. Attractor reconstruction. *Scholarpedia*, 1(10):1727, 2006.

[SBDH97] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke. Takens embedding theorems for forced and stochastic systems. *Nonlinear Anal.*, 30(8):5303–5314, 1997.

[Sem96] S. Semmes. On the nonexistence of bilipschitz parametrizations and geometric problems about A_∞-weights. *Rev. Mat. Iberoam.*, 12(2):337–410, 1996.

[She18] W. Shen. Hausdorff dimension of the graphs of the classical Weierstrass functions. *Math. Z.*, 289(1-2):223–266, 2018.

[Shm14] P. Shmerkin. On the exceptional set for absolute continuity of Bernoulli convolutions. *Geom. Funct. Anal.*, 24(3):946–958, 2014.

[Sim12] D. Simmons. Conditional measures and conditional expectation; Rohlin's disintegration theorem. *Discrete Contin. Dyn. Syst.*, 32(7):2565–2582, 2012.

[Sko08] A. Skopenkov. Embedding and knotting of manifolds in Euclidean spaces. In *Surveys in Contemporary Mathematics*, volume 347 of *London Math. Soc. Lecture Note Ser.*, pages 248–342. Cambridge Univ. Press, Cambridge, 2008. arXiv:math/0604045.

[SM90] G. Sugihara and R. M. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734–741, 1990.

[Sol95] B. Solomyak. On the random series $\sum \pm\lambda^n$ (an Erdős problem). *Ann. of Math. (2)*, 142(3):611–625, 1995.

[Sol98] B. Solomyak. Measure and dimension for some fractal families. *Math. Proc. Cambridge Philos. Soc.*, 124(3):531–546, 1998.

[Śpi25] A. Śpiewak. On the regularity of time-delayed embeddings with self-intersections. *Bull. Pol. Acad. Sci. Math.*, 73:47–68, 2025.

[SRAB17] D. Stotz, E. Riegler, E. Agustsson, and H. Bölcskei. Almost lossless analog signal separation and probabilistic uncertainty relations. *IEEE Trans. Inform. Theory*, 63(9):5445–5460, 2017.

[SS06] P. Shmerkin and B. Solomyak. Zeros of $\{-1, 0, 1\}$ power series and connectedness loci for self-affine sets. *Experiment. Math.*, 15(4):499–511, 2006.

[SSOY98] C. G. Schroer, T. Sauer, E. Ott, and J. A. Yorke. Predicting chaos most of the time from embeddings with self-intersections. *Phys. Rev. Lett.*, 80:1410–1413, 1998.

[SSU01a] K. Simon, B. Solomyak, and M. Urbański. Hausdorff dimension of limit sets for parabolic IFS with overlaps. *Pacific J. Math.*, 201(2):441–478, 2001.

[SSU01b] K. Simon, B. Solomyak, and M. Urbański. Invariant measures for parabolic IFS with overlaps and random continued fractions. *Trans. Amer. Math. Soc.*, 353(12):5145–5164, 2001.

[Sta99] J. Stark. Delay embeddings for forced systems. I. Deterministic forcing. *J. Nonlinear Sci.*, 9(3):255–332, 1999.

[SYC91] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *J. Stat. Phys.*, 65(3-4):579–616, 1991.

[Tak81] F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Math.*, pages 366–381. Springer, Berlin, 1981.

[Tak02] F. Takens. The reconstruction theorem for endomorphisms. *Bull. Braz. Math. Soc. (N.S.)*, 33(2):231–262, 2002.

[Var19] P. P. Varjú. On the dimension of Bernoulli convolutions for all transcendental parameters. *Ann. of Math. (2)*, 189(3):1001–1011, 2019.

[Whi36] H. Whitney. Differentiable manifolds. *Ann. of Math. (2)*, 37(3):645–680, 1936.

[Whi44] H. Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *Ann. of Math. (2)*, 45:220–246, 1944.

[WV10] Y. Wu and S. Verdú. Rényi information dimension: fundamental limits of almost lossless analog compression. *IEEE Trans. Inform. Theory*, 56(8):3721–3748, 2010.

[WV12] Y. Wu and S. Verdú. Optimal phase transitions in compressed sensing. *IEEE Trans. Inform. Theory*, 58(10):6241–6263, 2012.

[ZBD15] J. Zhu, D. Baron, and M. F. Duarte. Recovery from linear measurements with complexity-matching universal signal estimation. *IEEE Trans. Signal Process.*, 63(6):1512–1527, 2015.