Embedding theory in geometric measure theory

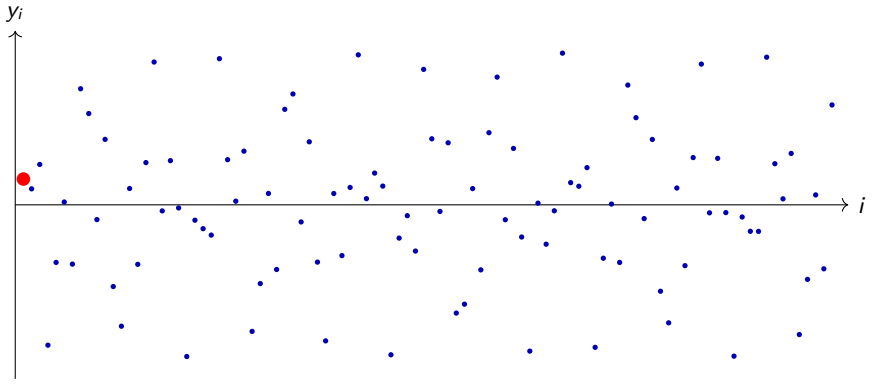
Part 3: Reconstruction of dynamics from data

Adam Śpiewak

Institute of Mathematics, Polish Academy of Sciences

Minicourse, Uppsala University

A scalar time series



The data was generated by a **deterministic dynamical system** $T\colon X \to X$ via a scalar observable $h\colon X \to \mathbb{R}$: $y_i = h(T^i x_0)$.

Task: given the past observations, **predict** the next value y_{i+1}.

Time-delayed embedding

Idea: pack k consecutive observations into a single higher-dimensional vector:

$$\mathbf{z}_i = (y_i,\ y_{i+1},\ \ldots,\ y_{i+k-1}) \in \mathbb{R}^k.$$

Hope: for large enough k, we might be able to find a reasonable **prediction map** $S\colon \mathbb{R}^k \to \mathbb{R}^k$, acting on the embedded time series as

$$(y_i,\, y_{i+1},\, \ldots,\, y_{i+k-1}) = z_i \xrightarrow{\quad S \quad} z_{i+1} = (y_{i+1},\, \ldots,\, y_{i+k}),$$

so that

$$z_{i+1} \simeq S(z_i)$$

If yes, we can treat $z_i \in \mathbb{R}^k$ as an orbit of S and predict (approximate) future values of the time series by iterating S.

Delay embedding with $k = 1$: no clear structure

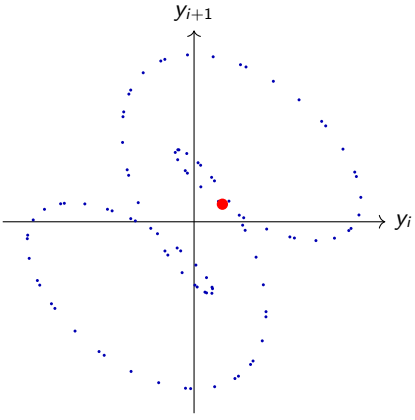


States with different dynamics can give the **same observation**.
Knowing y_i **does not determine** y_{i+1} uniquely.

No reasonable $S : \mathbb{R} \to \mathbb{R}$

Delay embedding with $k = 2$: dynamics unfolds!

(y_i, y_{i+1}) traces a curve — **dynamics (almost) reconstructed**.



Almost all points have a unique successor: $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ is almost well-defined.

The generating system

Dynamics: rotation on $\mathbb{S}^1$ by $\alpha = \frac{43\pi}{180}$

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\frac{43\pi}{180} & -\sin\frac{43\pi}{180} \\ \sin\frac{43\pi}{180} & \cos\frac{43\pi}{180} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Observable: $h(x, y) = 4\,x^3 - 2\,x + 0.35\,y$

The delay coordinate map

The mathematical study of time-delayed embeddings is based on the properties of the **delay coordinate map**:

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

If $\phi_{h,k}$ is injective on X: the state x is uniquely determined by k observations
$\Longrightarrow$ **prediction is possible** (even more: S is a faithful model of the original dynamics)

Central question: How large does k need to be?

Takens-type embedding theorems

Theorem (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.

Theorem (Sauer–Yorke–Casdagli 1991)**.** 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 $p = 1, \ldots, k-1$, where $\mathrm{Per}_p(T) = \{x \in X : T^p x = x\}$. Then for a prevalent Lipschitz observable $h \in \mathrm{Lip}(X, \mathbb{R})$, the delay coordinate map $\phi_{h,k}$ is injective on X.

Note: the threshold twice the dimension is the same as in Mañé's embedding theorem — two instances of the same transversality-based philosophy.

Smaller embedding dimension in practice

However: In numerical experiments, an embedding dimension k **significantly smaller** than $2\,\overline{\dim}_{\mathrm{B}}(X)$ often suffices for the task of predicting the dynamics.

Can we explain this rigorously?

Problem: as stated, Takens-type theorems are optimal. The conclusions are too strong!

Schroer, Sauer, Ott and Yorke (*Phys. Rev. Letters*, 1998) formulated two conjectures, proposing that the reduction of the embedding dimension should be possible in the probabilistic/statistical paradigm and $\mathbf{k} \sim \dim(\mathbf{X})$ should suffice.

The SSOY conjectures: predictability

Let $T\colon X \to X$ with a natural (physical) measure μ of information dimension $\mathrm{ID}(\mu)$, and let $y_i = h(T^i x)$.

Conjecture 1 (Predictability)**.** If $k > \mathrm{ID}(\mu)$, then for a typical observable h, the next value y_{i+k} is μ-**almost surely predictable** from $(y_i, \ldots, y_{i+k-1})$.

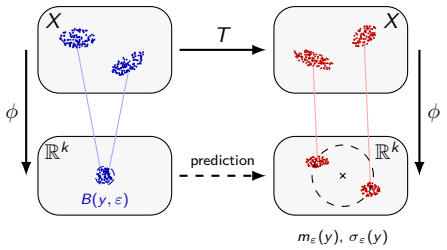
μ-almost surely predictable: there are a Borel set $X_h \subset X$ with $\mu(X_h) = 1$ and a prediction map $S\colon \phi_{h,k}(X_h) \to \mathbb{R}^k$ such that the diagram commutes on X_h:

$$\begin{array}{ccc} X_h & \xrightarrow{\ T\ } & X \\ {\scriptstyle \phi_{h,k}}\downarrow & & \downarrow{\scriptstyle \phi_{h,k}} \\ \phi_{h,k}(X_h) & \xrightarrow{\ S\ } & \mathbb{R}^k \end{array} \qquad \text{i.e.} \qquad \phi_{h,k} \circ T = S \circ \phi_{h,k} \ \text{ on } X_h.$$

Equivalently: $\phi_{h,k}(x_1) = \phi_{h,k}(x_2) \implies \phi_{h,k}(Tx_1) = \phi_{h,k}(Tx_2)$ for $x_1, x_2 \in X_h$.

The conjecture proposes **halving** the deterministic threshold $2\dim(X)$ in exchange for a probabilistic conclusion.

The SSOY conjectures: prediction error



$\phi = \phi_{h,k}$, $\mu_{y,\varepsilon} - \mu$ conditioned on $\phi^{-1}(B(y,\varepsilon))$:

$$m_\varepsilon(y) = \int \phi \circ T \, d\mu_{y,\varepsilon},$$

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

Prediction and prediction error: mean and standard deviation of $\phi(Tx)$, $x \sim \mu_{y,\varepsilon}$.

Conjecture 2 (Prediction error)**.** Let $D = \mathrm{ID}(\mu)$. For typical h, fixed $\delta > 0$ and small $\varepsilon > 0$:

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

(ii) if $D < k < 2D$, then $\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \asymp \varepsilon^{k-D}$ (power decay).

Results on the SSOY conjectures

Theorem (Barański–Gutman–Ś 2022–2026)**.** Let $X \subset \mathbb{R}^N$ be compact, $T\colon X \to X$ Lipschitz, μ a Borel probability measure on X and $k \in \mathbb{N}$. Then for a prevalent Lipschitz observable $h\colon X \to \mathbb{R}$, with $\phi = \phi_{h,k}$:

(i) if $k < \dim_{\mathrm{H}} \mu$ and μ is T-invariant and ergodic, then $\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \geq C > 0$ for small $\delta > 0$ and $0 < \varepsilon < \varepsilon_0(\delta)$ (prediction **fails**) [Conjecture 2(i)];

(ii) if $k > \dim_{\mathrm{H}} \mu$, then h is μ-almost surely predictable [Conjecture 1];

(iii) if $k > \overline{\dim}_{\mathrm{B}} X$, then for every $\delta, \theta > 0$: $\mu(\{x : \sigma_\varepsilon(\phi(x)) > \delta\}) \leq C\varepsilon^{k - \overline{\dim}_{\mathrm{B}} X - \theta}$ for small ε [Conjecture 2(ii)];

(iv) there is a C^∞ diffeomorphism T of $S^2 \times S^1$ with a natural measure μ, $\mathrm{ID}(\mu) < 1$, such that a prevalent h is not μ-a.s. 1-predictable [Conjectures 1 and 2(ii) fail for ID].

Probabilistic Takens theorem

We prove the most basic case of the probabilistic Takens theorem.

Theorem (Barański–Gutman–Ś 2020)**.** Let $X \subset \mathbb{R}^N$ be compact, $T\colon X \to X$ Lipschitz and injective, μ a Borel probability measure on X and $k \in \mathbb{N}$. Assume

$$k > \dim_{\mathrm{H}} \mu \qquad \text{and} \qquad \dim_{\mathrm{H}} \left(\mu|_{\mathrm{Per}_p(T)}\right) < p \quad \text{for } p = 1, \dots, k-1.$$

Then for a prevalent Lipschitz observable $h\colon X \to \mathbb{R}$, the delay coordinate map $\phi_{h,k}$ is μ-a.s. injective. Hence h is μ-almost surely predictable.

Polynomial perturbations

Fix a Lipschitz observable h and let $h_1, \dots, h_m$ be a basis of the polynomials on $\mathbb{R}^N$ of degree $\leq 2k-1$ (the *probe set*). For $\alpha \in \mathbb{R}^m$ let $h_\alpha = h + \sum_{j=1}^m \alpha_j h_j$.

By the definition of prevalence it suffices to prove the assertion for h_α, for Lebesgue-a.e. $\alpha \in B_m(0,1)$.

The perturbed delay map is **affine** in α: since $h_\alpha(T^i x) = h(T^i x) + \sum_{j=1}^m \alpha_j h_j(T^i x)$,

$$\phi_{h_\alpha,k}(x) = \begin{pmatrix} h_\alpha(x) \\ h_\alpha(Tx) \\ \vdots \\ h_\alpha(T^{k-1}x) \end{pmatrix} = \begin{pmatrix} h(x) \\ h(Tx) \\ \vdots \\ h(T^{k-1}x) \end{pmatrix} + \begin{pmatrix} h_1(x) & \cdots & h_m(x) \\ h_1(Tx) & \cdots & h_m(Tx) \\ \vdots & & \vdots \\ h_1(T^{k-1}x) & \cdots & h_m(T^{k-1}x) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}.$$

The perturbed delay map

Hence $\phi_{h_\alpha,k}(x) - \phi_{h_\alpha,k}(y) = D_{x,y}\,\alpha + w_{x,y}$ with $w_{x,y} = \phi_{h,k}(x) - \phi_{h,k}(y)$:

$$\phi_{h_\alpha,k}(x) - \phi_{h_\alpha,k}(y) = \underbrace{\begin{pmatrix} h_1(x) - h_1(y) & \cdots & h_m(x) - h_m(y) \\ \vdots & \ddots & \vdots \\ h_1(T^{k-1}x) - h_1(T^{k-1}y) & \cdots & h_m(T^{k-1}x) - h_m(T^{k-1}y) \end{pmatrix}}_{D_{x,y}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}$$

$$+ \underbrace{\begin{pmatrix} h(x) - h(y) \\ \vdots \\ h(T^{k-1}x) - h(T^{k-1}y) \end{pmatrix}}_{w_{x,y}},$$

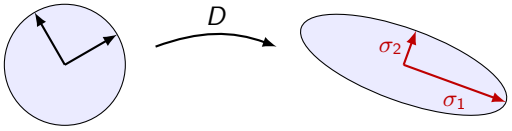
i.e. $D_{x,y} = \left[h_j(T^i x) - h_j(T^i y)\right]_{0 \le i < k,\ 1 \le j \le m}$.

Singular values

Definition

The *singular values* $\sigma_1(D) \geq \sigma_2(D) \geq \cdots \geq \sigma_{\min\{k,m\}}(D) \geq 0$ of a linear map $D\colon \mathbb{R}^m \to \mathbb{R}^k$ are the square roots of the $\min\{k, m\}$ largest eigenvalues of the symmetric matrix $D^T D$.

- D maps the unit ball $B_m(0,1)$ onto an ellipsoid with **semi-axes** $\sigma_1(D), \sigma_2(D), \ldots$



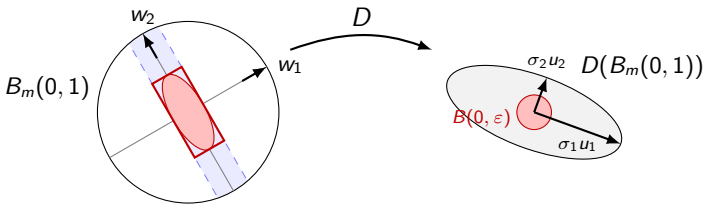
- $\sigma_1(D) = \|D\|$, and $\sigma_p(D) > 0 \iff \operatorname{rank} D \geq p$.

Transversality inequality

Lemma

Let $D\colon \mathbb{R}^m \to \mathbb{R}^k$ be linear with $\sigma_p(D) > 0$. Then for $z \in \mathbb{R}^k$ and $\varepsilon > 0$,

$$\mathrm{Leb}_m\big(\{\alpha \in B_m(0,1) : |D\alpha + z| \le \varepsilon\}\big) \le C_{m,k}\Big(\frac{\varepsilon}{\sigma_p(D)}\Big)^p.$$



Rank of the observation matrices

$\mathrm{Orb}(x) = \{y \in X : T^n y = x \text{ or } T^n x = y \text{ for some } n \geq 0\}$ – the full orbit of x.

Lemma (Rank lemma)

Let T be injective and let $h_1, \ldots, h_m$ be a basis of the polynomials on $\mathbb{R}^N$ of degree $\leq 2k-1$. Then:

1. *if $y \neq x$, then* $\operatorname{rank} D_{x,y} \geq 1$;
2. *if $y \notin \mathrm{Orb}(x)$ and y has minimal period $p \in \{1, \ldots, k-1\}$, then* $\operatorname{rank} D_{x,y} \geq p$;
3. *if $y \notin \mathrm{Orb}(x)$ and y is not periodic with minimal period smaller than k, then* $\operatorname{rank} D_{x,y} = k$.

- Periodic points cause a **rank drop** – hence the assumption $\dim_{\mathrm{H}}(\mu|_{\mathrm{Per}_p(T)}) < p$.

Interpolation

Lemma

Let $h_1, \ldots, h_m$ be a basis of the polynomials on $\mathbb{R}^N$ of degree $\leq 2k-1$. If $z_1, \ldots, z_s \in \mathbb{R}^N$ ***are distinct*** *and $s \leq 2k$, then the linear map $V \colon \mathbb{R}^m \to \mathbb{R}^s$,*

$$V = \begin{pmatrix} h_1(z_1) & \cdots & h_m(z_1) \\ \vdots & & \vdots \\ h_1(z_s) & \cdots & h_m(z_s) \end{pmatrix}$$

has full rank s, i.e. V is onto.

- **Key fact:** polynomials of degree $\leq 2k-1$ **interpolate** arbitrary values at $\leq 2k$ distinct points: for distinct $z_1, \ldots, z_s$, $s \leq 2k$, and any $y_1, \ldots, y_s \in \mathbb{R}$ there is a polynomial P of degree $\leq 2k-1$ with $P(z_l) = y_l$ for $l = 1, \ldots, s$.
- Take any $y = (y_1, \ldots, y_s) \in \mathbb{R}^s$ and write the interpolating polynomial in the basis as $P = \sum_{j=1}^{m} \alpha_j h_j$. Then $V\alpha = \big(P(z_1), \ldots, P(z_s)\big) = y$, so V is onto.

The factorization $D_{x,y} = JV$

Let $z_1, \dots, z_s$ be the distinct points among $x, Tx, \dots, T^{k-1}x, y, Ty, \dots, T^{k-1}y$, so $s \le 2k$.

- $V = \big[h_j(z_l)\big]_{l \le s,\, j \le m} \colon \mathbb{R}^m \to \mathbb{R}^s$ has rank s by interpolation, i.e. V is onto;
- $J = J_{x,y} \colon \mathbb{R}^s \to \mathbb{R}^k$ has entries $J_{il} = \mathbb{1}_{\{T^i x\}}(z_l) - \mathbb{1}_{\{T^i y\}}(z_l)$, $0 \le i < k$, $1 \le l \le s$.

Then

$$D_{x,y} = JV \qquad \text{and} \qquad \operatorname{rank} D_{x,y} = \operatorname{rank} J.$$

Example ($k = 2$, x, Tx, y, Ty distinct):

$$D_{x,y} = \underbrace{\begin{array}{c} \begin{array}{cccc} x & Tx & y & Ty \end{array} \\ \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \end{array}}_{J} \underbrace{\begin{pmatrix} h_1(x) & \cdots & h_m(x) \\ h_1(Tx) & \cdots & h_m(Tx) \\ h_1(y) & \cdots & h_m(y) \\ h_1(Ty) & \cdots & h_m(Ty) \end{pmatrix}}_{V}, \qquad \operatorname{rank} J = 2 = k.$$

The factorization $D_{x,y} = JV$: periodic points

Example ($k = 3$): x, y, Ty distinct and $x = Tx = T^2x$, $T^2y = y$. Then $s = 3$, $(z_1, z_2, z_3) = (x, y, Ty)$ and

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

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

The first and the third row of J coincide, so $\operatorname{rank} D_{x,y} = \operatorname{rank} J = 2 < k$ – the rank drops at periodic points (Rank lemma (ii) with $p = 2$).

Rank lemma (i)

Claim (i). If $y \neq x$, then $\operatorname{rank} D_{x,y} \geq 1$.

Proof. As $x \neq y$, the points x and y correspond to different columns, so

$$J_{x,y} = \begin{pmatrix} x & & y & \\ 1 & 0 \cdots 0 & -1 & 0 \cdots 0 \\ * & * & * & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & * & * \end{pmatrix}.$$

The first row is non-zero, hence $\operatorname{rank} D_{x,y} = \operatorname{rank} J_{x,y} \geq 1$.

Rank lemma (ii)

Claim (ii). If $y \notin \mathrm{Orb}(x)$ and y **has minimal period** $p \in \{1, \dots, k-1\}$, then rank $D_{x,y} \geq p$.

Proof.

- by injectivity and $y \notin \mathrm{Orb}(x)$, we have $\mathrm{Orb}(x) \cap \mathrm{Orb}(y) = \emptyset$, so iterates of x and y give different columns
- $y, Ty, \dots, T^{p-1}y$ are distinct

$$J_{x,y} = \left(\begin{array}{cccccc} x & \cdots & y & Ty & \cdots & T^{p-1}y \\ * & \cdots & -1 & 0 & \cdots & 0 \\ * & \cdots & 0 & -1 & \cdots & 0 \\ \vdots & & \vdots & & \ddots & \vdots \\ * & \cdots & 0 & 0 & \cdots & -1 \\ \vdots & & \vdots & & & \vdots \end{array}\right).$$

The rows $i = 0, \dots, p-1$ and the columns $y, \dots, T^{p-1}y$ form $-\mathrm{Id}_p$, hence rank $D_{x,y} =$ rank $J_{x,y} \geq p$.

Rank lemma (iii)

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

Proof.

- as before, the points $T^i x$ and $T^j y$ give different columns;
- $y, Ty, \ldots, T^{k-1}y$ are distinct

Hence

$$J_{x,y} = \left(\begin{array}{ccc|ccc} x & \cdots & y & Ty & \cdots & T^{k-1}y \\ * & \cdots & -1 & 0 & \cdots & 0 \\ * & \cdots & 0 & -1 & \cdots & 0 \\ \vdots & & \vdots & & \ddots & \vdots \\ * & \cdots & 0 & 0 & \cdots & -1 \end{array}\right).$$

The columns $y, \ldots, T^{k-1}y$ form $-\mathrm{Id}_k$, hence rank $D_{x,y}$ = rank $J_{x,y} = k$.

Sketch of proof: reduction

Write $\phi_\alpha = \phi_{h_\alpha,k}$ for $\alpha \in B_m(0,1)$.

- **Good sets.** As $\dim_{\mathrm{H}} \mu < k$ and $\dim_{\mathrm{H}}(\mu|_{\mathrm{Per}_p(T)}) < p$, we can choose Borel sets

$$Y_p \subset \mathrm{Per}_p(T) \setminus \bigcup_{q<p} \mathrm{Per}_q(T), \quad \mathcal{H}^p(Y_p) = 0 \qquad \text{for } p = 1, \dots, k-1,$$
$$Y_k \subset X \setminus \bigcup_{q<k} \mathrm{Per}_q(T), \quad \mathcal{H}^k(Y_k) = 0,$$

  such that $Y = Y_1 \cup \cdots \cup Y_k$ satisfies $\mu(Y) = 1$.

- **Fubini.** As in Mañé's theorem, it suffices to show that for every $x \in Y$

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

- **Affine structure.** $\phi_\alpha(x) = \phi_\alpha(y) \iff D_{x,y}\alpha + w_{x,y} = 0.$

Sketch of proof: splitting B_x

Fix $x \in Y$. Let $Y_0 = \mathrm{Orb}(x) \setminus \{x\}$ and replace Y_p by $Y_p \setminus \mathrm{Orb}(x)$ for $p = 1, \ldots, k$. Then

$$Y \setminus \{x\} \subset Y_0 \cup Y_1 \cup \cdots \cup Y_k.$$

Splitting according to where y with $\phi_\alpha(x) = \phi_\alpha(y)$ belongs,

$$B_x \subset \bigcup_{p=0}^{k} B_{x,p}, \qquad \text{where} \quad B_{x,p} = \{\alpha \in B_m(0,1) : \phi_\alpha(x) = \phi_\alpha(y) \text{ for some } y \in Y_p\}.$$

The set $B_{x,0}$. For $y \in Y_0$ we have $x \neq y$ so by Rank lemma $D_{x,y} \neq 0$ and

$$\{\alpha : D_{x,y}\alpha + w_{x,y} = 0\}$$

is a proper affine subspace of $\mathbb{R}^m$ (or empty). As $Y_0 \subset \mathrm{Orb}(x)$ is countable, $\mathrm{Leb}_m(B_{x,0}) = 0$.

Sketch of proof: the sets $B_{x,p}$, $p \geq 1$

By the Rank lemma $\sigma_p(D_{x,y}) > 0$ for $y \in Y_p$. Hence $B_{x,p} = \bigcup_{n=1}^{\infty} B_{x,p,n}$, where

$$B_{x,p,n} = \{\alpha \in B_m(0,1) : \phi_\alpha(x) = \phi_\alpha(y) \text{ for some } y \in Y_p \text{ with } \sigma_p(D_{x,y}) \geq 1/n\}.$$

As before, the transversality inequality and covering Y_p (with $\mathcal{H}^p(Y_p) = 0$) give $\mathrm{Leb}_m(B_{x,p,n}) = 0$. Hence $\mathrm{Leb}_m(B_x) = 0$. □

Beyond injectivity: predictability

To deal with the (quantitative) **predictability** required by the SSOY conjectures (rather than injectivity), one proves a much more involved version of the Rank lemma.

Proposition (Barański–Gutman–Ś 2024)**.** For every $T\colon X \to X$ on a set $X \subset \mathbb{R}^N$ and every $x, y \in X$ at least one of the following holds:

(i) $\operatorname{rank} D_{x,y} = k$;

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