Learning theory for dynamics

Nature, limits and accuracy of learning, in the context of a dynamical system

Suddhasattwa Das

Department of Mathematics and Statistics
Texas Tech University

iamsuddhasattwa@gmail.com

February 02, 2023

What is learning

Given a set of inputs $x_1, x_2, \ldots \in U \subseteq \mathbb{R}^D$, and outputs $y_1, y_2, \ldots \in V \subseteq \mathbb{R}^d$, a learning task is to find a function

$$f : U \to V, \quad f(x_n) = y_n, \quad n = 1, 2, 3, \ldots.$$

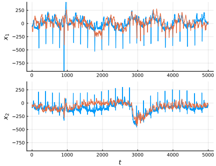
f is usually searched for from a hypothesis / search space $\mathcal{H}$.

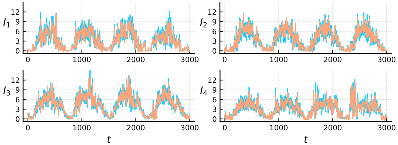
What is learning

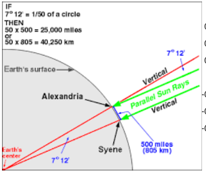
Given a set of inputs $x_1, x_2, \ldots \in U \subseteq \mathbb{R}^D$, and outputs $y_1, y_2, \ldots \in V \subseteq \mathbb{R}^d$, a learning task is to find a function

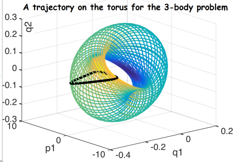
$$f : U \to V, \quad f(x_n) = y_n, \quad n = 1, 2, 3, \ldots.$$

f is usually searched for from a hypothesis / search space $\mathcal{H}$.

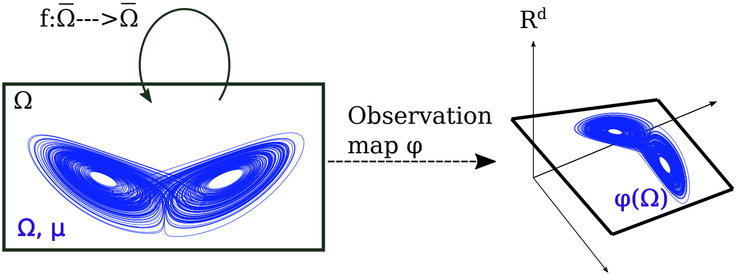




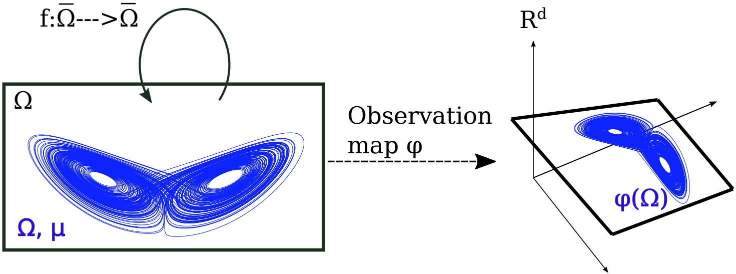




General dynamics

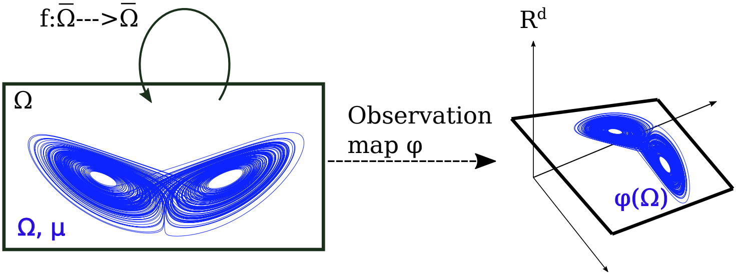


General dynamics



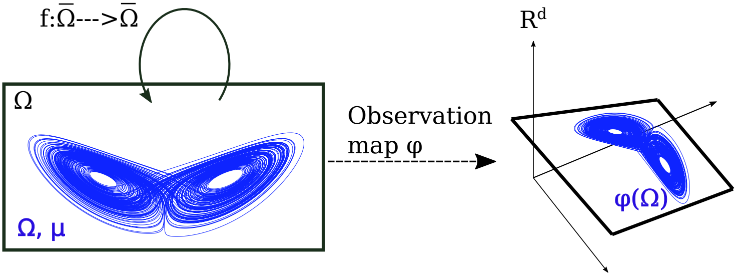
- A *dynamical system* $f : \tilde{\Omega} \to \tilde{\Omega}$ on a manifold $\tilde{\Omega}$.

General dynamics



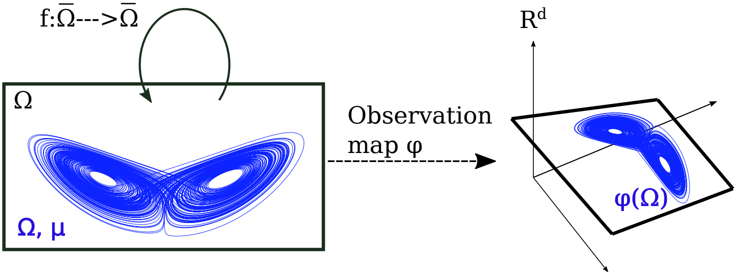
- A *dynamical system* $f : \tilde{\Omega} \to \tilde{\Omega}$ on a manifold $\tilde{\Omega}$.
- An invariant ergodic measure μ, with compact support Ω.

General dynamics



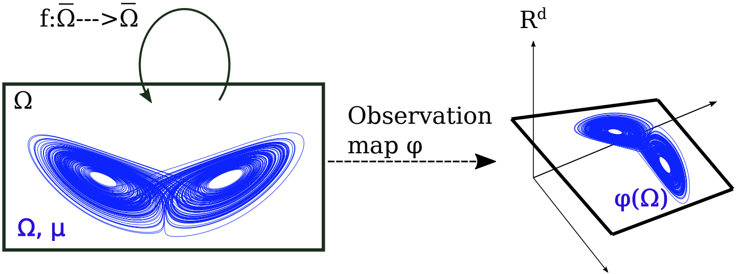
- A *dynamical system* $f : \tilde{\Omega} \to \tilde{\Omega}$ on a manifold $\tilde{\Omega}$.
- An invariant ergodic measure μ, with compact support Ω.
- A C^r function $\phi : \Omega \to \mathbb{R}^d$.

General dynamics



- A *dynamical system* $f : \tilde{\Omega} \to \tilde{\Omega}$ on a manifold $\tilde{\Omega}$.
- An invariant ergodic measure μ, with compact support Ω.
- A C^r function $\phi : \Omega \to \mathbb{R}^d$.
- An orbit $\omega_0, \omega_1, \omega_2, \ldots,$, where $\omega_n = f^n(\omega_0)$.

General dynamics



- A *dynamical system* $f : \tilde{\Omega} \to \tilde{\Omega}$ on a manifold $\tilde{\Omega}$.
- An invariant ergodic measure μ, with compact support Ω.
- A C^r function $\phi : \Omega \to \mathbb{R}^d$.
- An orbit $\omega_0, \omega_1, \omega_2, \ldots,$, where $\omega_n = f^n(\omega_0)$.
- A timeseries $y_n = \phi(\omega_n)$, $n = 1, 2, 3, \ldots$.

Learning dynamics

Find a dynamical system $\hat{f} : \mathbb{R}^D \to \mathbb{R}^D$ on some Euclidean space $\mathbb{R}^D$ with the dynamics,

$$\hat{x}_{n+1} = \hat{f}\left(\hat{x}_n\right), \quad n = 1, 2, 3, \ldots,$$

and an observation map $\hat{\phi} : \mathbb{R}^D \to \mathbb{R}^d$ such that

$$\phi(\omega_n) = y_n = \hat{\phi}\left(\hat{x}_n\right), \quad n = 1, 2, 3, \ldots.$$

The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

The Koopman operator U^t

Dynamics / trajectories on phase space ↔ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

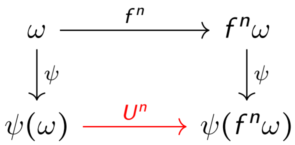
$$(U^n \psi)(x) = \psi(f^n x).$$

The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

$$(U^n \psi)(x) = \psi\left(f^n x\right).$$



The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

$$(U^n \psi)(x) = \psi\left(f^n x\right).$$

$$\begin{array}{ccc} \omega & \xrightarrow{\quad f^n \quad} & f^n \omega \\ \downarrow \psi & & \downarrow \psi \\ \psi(\omega) & \xrightarrow{\quad U^n \quad} & \psi(f^n \omega) \end{array}$$

Space of observables :

.

The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

$$(U^n \psi)(x) = \psi\left(f^n x\right).$$

$$\begin{array}{ccc} \omega & \xrightarrow{\quad f^n \quad} & f^n\omega \\ \big\downarrow \psi & & \big\downarrow \psi \\ \psi(\omega) & \xrightarrow{\quad U^n \quad} & \psi(f^n\omega) \end{array}$$

Space of observables : $\quad M \to \mathbb{R}$

.

The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

$$(U^n \psi)(x) = \psi\left(f^n x\right).$$

$$\begin{array}{ccc} \omega & \xrightarrow{\quad f^n \quad} & f^n \omega \\ \downarrow \psi & & \downarrow \psi \\ \psi(\omega) & \xrightarrow{\quad U^n \quad} & \psi(f^n \omega) \end{array}$$

Space of observables : $M \to \mathbb{R}$, $C^0(M)$

.

The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

$$(U^n \psi)(x) = \psi\,(f^n x)\,.$$

$$\begin{array}{ccc}
\omega & \xrightarrow{\;f^n\;} & f^n\omega \\
\downarrow \psi & & \downarrow \psi \\
\psi(\omega) & \xrightarrow{\;U^n\;} & \psi(f^n\omega)
\end{array}$$

Space of observables : $M \to \mathbb{R}$, $C^0(M)$, $C^r(M)$

.

The Koopman operator U^t

Dynamics / trajectories on phase space $\leftrightarrow$ Dynamics in spaces of observables.

Given a function $\psi : \Omega \to \mathbb{R}$, $(U^n \psi)$ is a new function defined as

$$(U^n \psi)(x) = \psi(f^n x) .$$

$$\begin{array}{ccc} \omega & \xrightarrow{\ f^n\ } & f^n \omega \\ \downarrow \psi & & \downarrow \psi \\ \psi(\omega) & \xrightarrow{\ U^n\ } & \psi(f^n \omega) \end{array}$$

Space of observables : $M \to \mathbb{R}$, $C^0(M)$, $C^r(M)$ or $L^2(\mu)$.

Embedding

There is an injective map $\Phi : \Omega \to \mathbb{R}^L$, and a map $g : \mathbb{R}^d \times \mathbb{R}^L \to \mathbb{R}^L$ such that

$$\Phi \circ f = g \circ (\phi \times \Phi)$$

The map ϕ is the measurement through which the dynamical system is observed. So the codomain of ϕ is often low dimensional and may be only partially observe the system. Since Φ is an embedding, it effectively serves as a representation of the dynamics-space Ω in $\mathbb{R}^L$ space. The function g connects the dynamics, with the embedding and the measurement.

Paradigm I - Delay embeddings $[\Phi \circ f = g \circ (\phi \times \Phi)]$

$$\{y_1, y_2, , \dots, y_n, \dots\} \mapsto \left\{ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{1+Q-1} \end{bmatrix}, \begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ y_{2+Q-1} \end{bmatrix}, \dots \begin{bmatrix} y_n \\ y_{n+1} \\ \vdots \\ y_{n+Q-1} \end{bmatrix}, \dots \right\},$$

Set $L = Qd$.

$$\Phi : \Omega \to \mathbb{R}^L, \ \Phi : \omega \mapsto \begin{bmatrix} \phi(\omega) \\ \vdots \\ \phi\left(f^{Q-1}\omega\right) \end{bmatrix}, \quad y_n^{(Q)} = \begin{bmatrix} y_n \\ y_{n+1} \\ \vdots \\ y_{n+Q-1} \end{bmatrix} = \Phi(f^n \omega).$$

$$g : \mathbb{R}^d \times \mathbb{R}^L \to \mathbb{R}^L, \ g : u \times \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_Q \end{bmatrix} \mapsto \begin{bmatrix} u \\ z_1 \\ \vdots \\ z_{Q-1} \end{bmatrix}$$

Paradigm II - Reservoir computing $[\Phi \circ f = g \circ (\phi \times \Phi)]$

A particular instance of g is

$$g(u, y) = \tanh\left(W_{in} u + W_Y y + v_{bias}\right),$$

where W_{in}, W_Y are random matrices of dimensions $L \times d$, $L \times L$ respectively, v_{bias} is a random vector of dimension L, and $\|W_Y\| \le \lambda < 1$. Using this g one can build a *reservoir* system, which is a skew product system on $\Omega \times \mathbb{R}^L$ defined as

$$\begin{pmatrix} \omega_{n+1} \\ y_{n+1} \end{pmatrix} := T_{\text{reservoir}} \begin{pmatrix} \omega_n \\ y_n \end{pmatrix} := \begin{pmatrix} f(\omega_n) \\ g\left(\phi(\omega), y_n\right) \end{pmatrix}. \qquad (1)$$

The paradigm of invariant graphs has been used in reservoir computing. It is popular due to the simplicity of its construction, and ease of use in learning problems. They are known for their robust performance in prediction but also for recovering other properties such as Lyapunov exponents.

The feedback function

Since Φ is an embedding, it effectively serves as a representation of the dynamics-space Ω in $\mathbb{R}^L$ space. The function g is explicitly known and computable. Note that $\Phi \circ f$ is the evolution of Φ under one iteration of the dynamics of f. Thus g contains and encodes the evolution law, in terms of the current states of Φ and ϕ.

The feedback function

Since Φ is an embedding, it effectively serves as a representation of the dynamics-space Ω in $\mathbb{R}^L$ space. The function g is explicitly known and computable. Note that $\Phi \circ f$ is the evolution of Φ under one iteration of the dynamics of f. Thus g contains and encodes the evolution law, in terms of the current states of Φ and ϕ.
For every $k \in \mathbb{N}$, there is a function :

$$w_k : \mathbb{R}^L \to \mathbb{R}^d, \quad w_k \circ \Phi = U^k \phi = \phi \circ f^k. \tag{2}$$

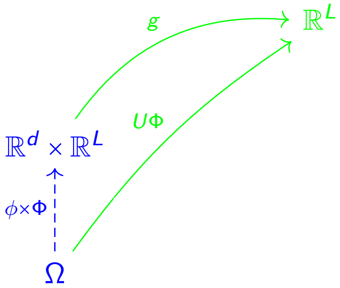
The feedback function

Since Φ is an embedding, it effectively serves as a representation of the dynamics-space Ω in $\mathbb{R}^L$ space. The function g is explicitly known and computable. Note that $\Phi \circ f$ is the evolution of Φ under one iteration of the dynamics of f. Thus g contains and encodes the evolution law, in terms of the current states of Φ and ϕ.
For every $k \in \mathbb{N}$, there is a function :

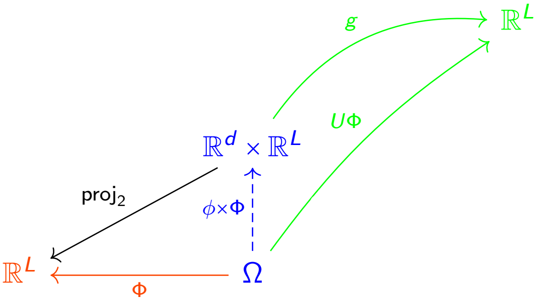
$$w_k : \mathbb{R}^L \to \mathbb{R}^d, \quad w_k \circ \Phi = U^k \phi = \phi \circ f^k. \tag{2}$$

Learning dynamics is about learning w_k.

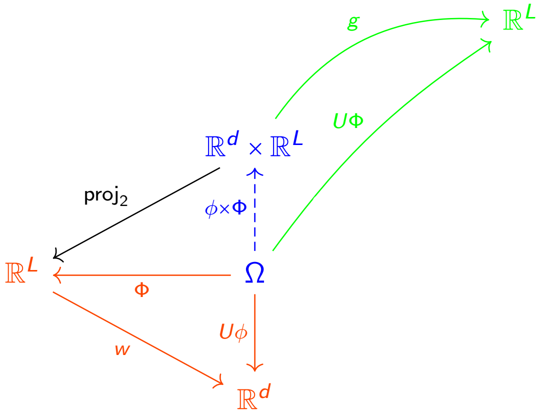
Maps and functions - IIa



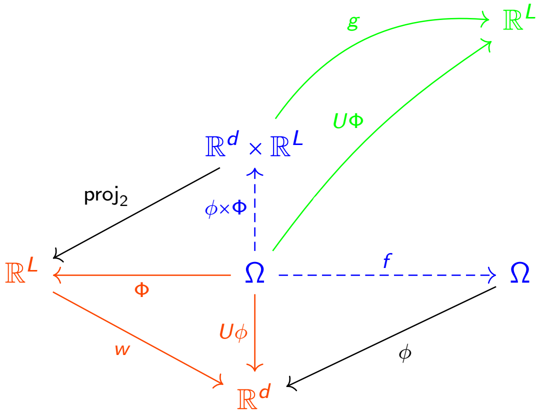
Maps and functions - IIb



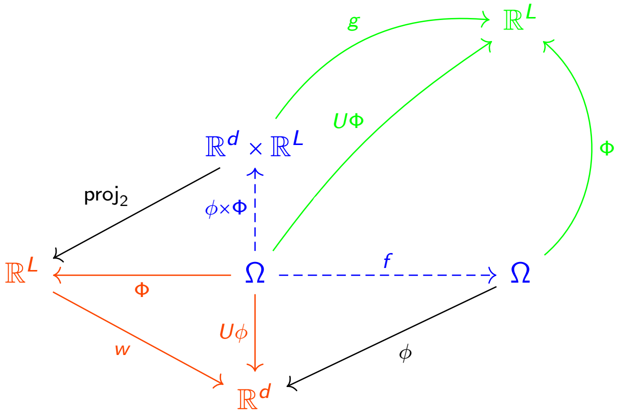
Maps and functions - IIc



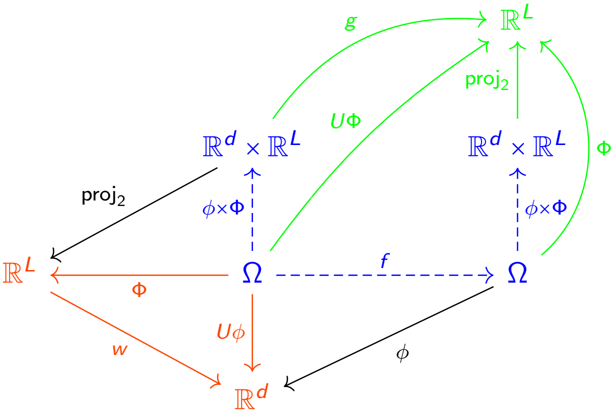
Maps and functions - IId



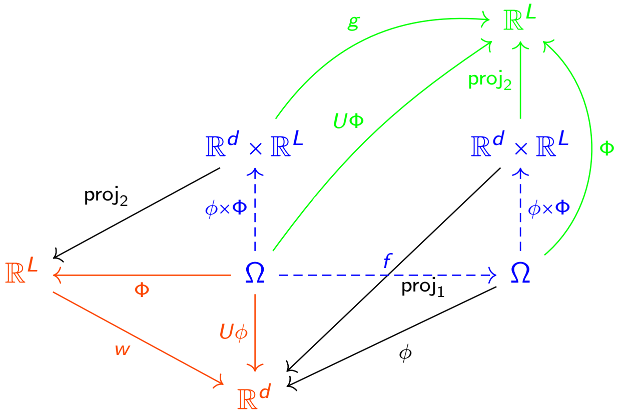
Maps and functions - IIe



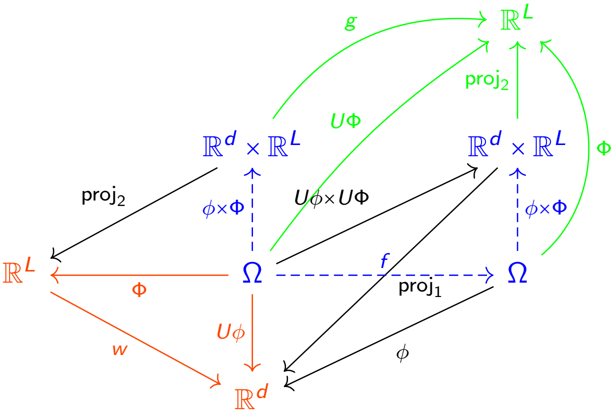
Maps and functions - IIf



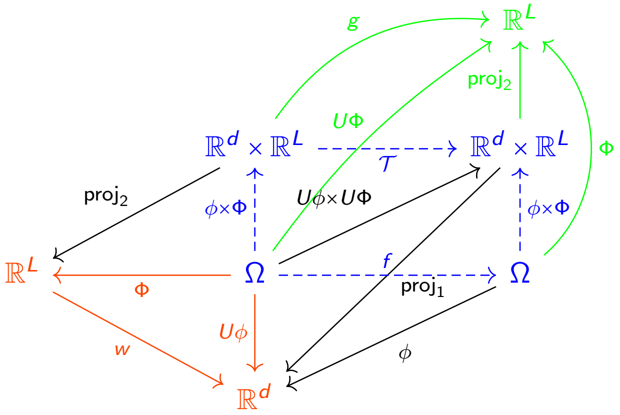
Maps and functions - IIg



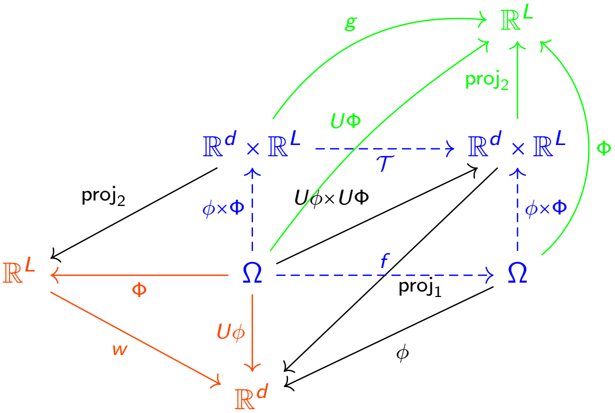
Maps and functions - IIh



Maps and functions - II



The reconstruction : $\mathcal{T} = (w \circ \mathrm{proj}_2) \times g$



$$\mathcal{T} : \mathbb{R}^d \times \mathbb{R}^L \to \mathbb{R}^d \times \mathbb{R}^L, \quad \begin{bmatrix} u_{n+1} \\ y_{n+1} \end{bmatrix} = \mathcal{T} \begin{bmatrix} u_n \\ y_n \end{bmatrix} = \begin{bmatrix} w(y_n) \\ g(u_n, y_n) \end{bmatrix}.$$

Overview

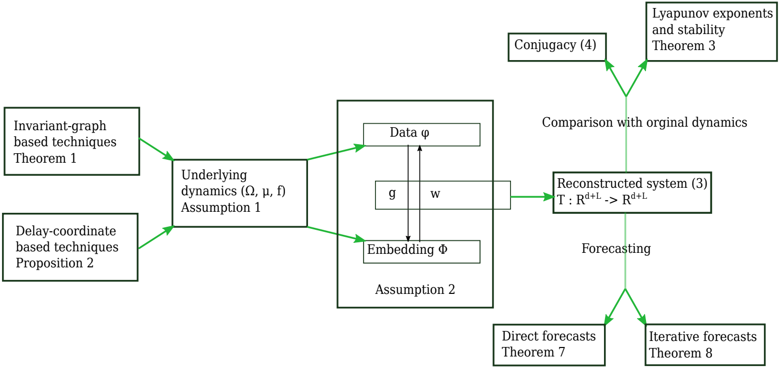


Figure: Outline of results and theory.

Lyapunov exponents : $F : M^m \to M^m$

Given a point (x, v) in TM, consider the limit

$$\lambda(x, v) := \limsup_{n\to\infty} \frac{1}{n} \ln \|DF^n(x)v\|$$

$\lambda(x, v)$ measures the *asymptotic exponential rate* of growth of perturbations at the point x, and in the direction v.

Lyapunov exponents : $F : M^m \to M^m$

Given a point (x, v) in TM, consider the limit

$$\lambda(x, v) := \limsup_{n \to \infty} \frac{1}{n} \ln \|DF^n(x)v\|$$

$\lambda(x, v)$ measures the *asymptotic exponential rate* of growth of perturbations at the point x, and in the direction v.

- Oseledet, Raghunathan, Froyland et. al. proved that with probability 1, the $\lambda(x, v)$ takes values from a finite collection of numbers, known as Lyapunov exponents

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m.$$

Lyapunov exponents : $F : M^m \to M^m$

Given a point (x, v) in TM, consider the limit

$$\lambda(x, v) := \limsup_{n\to\infty} \frac{1}{n} \ln \|DF^n(x)v\|$$

$\lambda(x, v)$ measures the *asymptotic exponential rate* of growth of perturbations at the point x, and in the direction v.

- Oseledet, Raghunathan, Froyland et. al. proved that with probability 1, the $\lambda(x, v)$ takes values from a finite collection of numbers, known as Lyapunov exponents

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m.$$

- *Nonuniform hyperbolicity theory* [LS. Young, Ya. Pesin, A. Katok et. al.] : over a set of full measure, the asymptotic behavior of the dynamics splits the tangent bundle into m line bundles, each representing a degree of stability / instability. Each of these sub-bundles contribute to stable and unstable manifolds.

Spurious Lyapunov exponents

The dynamics $f : \Omega \to \Omega$ has m Lyapunov exponents, where $m = \dim\left(\tilde{\Omega}\right)$.

$$
\begin{array}{ccccccc}
\tilde{\Omega} & \xleftarrow{\subseteq} & \Omega & \xrightarrow[h]{\phi\times\Phi} & (\phi\times\Phi)(\Omega) & \xrightarrow{\subset} & \mathbb{R}^{d+L} \\
\downarrow f & & \downarrow f & & \downarrow \mathcal{T} & & \downarrow \mathcal{T} \\
\tilde{\Omega} & \xleftarrow[\subseteq]{} & \Omega & \xrightarrow[\phi\times\Phi]{} & (\phi\times\Phi)(\Omega) & \xrightarrow[\subset]{} & \mathbb{R}^{d+L}
\end{array}
$$

Spurious Lyapunov exponents

The dynamics $f : \Omega \to \Omega$ has m Lyapunov exponents, where $m = \dim\left(\tilde{\Omega}\right)$.

$$
\begin{array}{ccccccc}
\tilde{\Omega} & \xleftarrow{\subseteq} & \Omega & \xrightarrow[h]{\phi\times\Phi} & (\phi \times \Phi)(\Omega) & \xrightarrow{\subset} & \mathbb{R}^{d+L} \\
\downarrow f & & \downarrow f & & \downarrow \mathcal{T} & & \downarrow \mathcal{T} \\
\tilde{\Omega} & \xleftarrow[\subseteq]{} & \Omega & \xrightarrow[\phi\times\Phi]{} & (\phi \times \Phi)(\Omega) & \xrightarrow[\subset]{} & \mathbb{R}^{d+L}
\end{array}
$$

- These m Lyapunov exponents are a subset of the $d + L$ Lyapunov exponents of $\mathcal{T}$ on the isomorphic image $h(\Omega)$.

Spurious Lyapunov exponents

The dynamics $f : \Omega \to \Omega$ has m Lyapunov exponents, where $m = \dim\left(\tilde{\Omega}\right)$.

$$
\begin{array}{ccccccc}
\tilde{\Omega} & \xleftarrow{\subseteq} & \Omega & \xrightarrow[h]{\phi\times\Phi} & (\phi\times\Phi)(\Omega) & \xrightarrow{\subset} & \mathbb{R}^{d+L} \\
\downarrow f & & \downarrow f & & \downarrow \mathcal{T} & & \downarrow \mathcal{T} \\
\tilde{\Omega} & \xleftarrow[\subseteq]{} & \Omega & \xrightarrow[\phi\times\Phi]{} & (\phi\times\Phi)(\Omega) & \xrightarrow[\subset]{} & \mathbb{R}^{d+L}
\end{array}
$$

- These m Lyapunov exponents are a subset of the $d + L$ Lyapunov exponents of $\mathcal{T}$ on the isomorphic image $h(\Omega)$.
- Since the original dynamics is being embedded in higher dimensional Euclidean space, an extra $d + L - m$ Lyapunov exponents are created.
- This could lead to additional instabilities.

Non-uniqueness of the learning target

Lyapunov exponents $\lambda_i(\mathcal{T})$ depend not only on the invariant set $X = h(\Omega)$ but also on its neighborhood. The function $w : \mathbb{R}^L \to \mathbb{R}^d$ is defined uniquely only on X and can be arbitrarily extended.

$$\mathfrak{S} := \left\{ \hat{w} \in C^1 \left(\mathbb{R}^L ; \mathbb{R}^d \right) \ : \ \hat{w}|_X = w|_X \right\},$$

Every $\hat{w} \in \mathfrak{S}$ is a C^1 function satisfying $\hat{w} \circ \Phi(\omega) = (U\phi)(\omega)$ for every $\omega \in \Omega$. Thus the target learning function is not precise, it is any function from the collection $\mathfrak{S}$.

Non-uniqueness of the learning target

Lyapunov exponents $\lambda_i(\mathcal{T})$ depend not only on the invariant set $X = h(\Omega)$ but also on its neighborhood. The function $w : \mathbb{R}^L \to \mathbb{R}^d$ is defined uniquely only on X and can be arbitrarily extended.

$$\mathfrak{S} := \left\{ \hat{w} \in C^1\left(\mathbb{R}^L; \mathbb{R}^d\right) \;:\; \hat{w}|_X = w|_X \right\},$$

Every $\hat{w} \in \mathfrak{S}$ is a C^1 function satisfying $\hat{w} \circ \Phi(\omega) = (U\phi)(\omega)$ for every $\omega \in \Omega$. Thus the target learning function is not precise, it is any function from the collection $\mathfrak{S}$.

The top Lyapunov exponent $\lambda_1(\mathcal{T})$ depends continuously on $\mathfrak{S}$:

$$\lambda_1 : \mathfrak{S} \to \mathbb{R}, \quad \lambda_1(\bar{w}) := \lambda_1(\mathcal{T}).$$

Non-uniqueness of the learning target

Lyapunov exponents $\lambda_i(\mathcal{T})$ depend not only on the invariant set $X = h(\Omega)$ but also on its neighborhood. The function $w : \mathbb{R}^L \to \mathbb{R}^d$ is defined uniquely only on X and can be arbitrarily extended.

$$\mathfrak{S} := \left\{ \hat{w} \in C^1\left(\mathbb{R}^L; \mathbb{R}^d\right) \ : \ \hat{w}|_X = w|_X \right\},$$

Every $\hat{w} \in \mathfrak{S}$ is a C^1 function satisfying $\hat{w} \circ \Phi(\omega) = (U\phi)(\omega)$ for every $\omega \in \Omega$. Thus the target learning function is not precise, it is any function from the collection $\mathfrak{S}$.

The top Lyapunov exponent $\lambda_1(\mathcal{T})$ depends continuously on $\mathfrak{S}$:

$$\lambda_1 : \mathfrak{S} \to \mathbb{R}, \quad \lambda_1(\bar{w}) := \lambda_1(\mathcal{T}).$$

This leads to the *structural constant*

$$\text{stability gap} := \inf_{\bar{w} \in \mathfrak{S}} \lambda_1(\bar{w}) - \lambda_1(f, \mu).$$

This is an outcome of the dynamics f and embedding mechanism $g : \mathbb{R}^{d+L} \to \mathbb{R}^L$.

Theorem 1. Stability gap

Assumption 3 The embedding mechanism g satisfies

$$\sup_{\omega\in\Omega} \|\partial_1 g\| |_{(\phi(\omega),\Phi(\omega))} \le 1, \quad \sup_{\omega\in\Omega} \|\partial_2 g\| |_{(\phi(\omega),\Phi(\omega))} \le 1.$$

Assumption 4 There is a continuous retraction $\mathrm{ret} : \mathcal{U} \to \mathrm{ran}\,\Phi$, for some open neighborhood $\mathcal{U}$ of $\mathrm{ran}\,\Phi$ in $\mathbb{R}^L$.

$$\kappa_{\mathrm{ret}} := \sup_{y\in \mathrm{ran}\,\Phi} \limsup_{y'\to y} \frac{d\left(\mathrm{ret}(y), \mathrm{ret}(y')\right)}{d\left(y, y'\right)}.$$

$$C_{\phi,\Phi} : \Omega \to \mathbb{R}^+, \quad C_{\phi,\Phi}(\omega) := \sup\left\{ \frac{\|D\phi(\omega)v\|}{\|D\Phi(\omega)v\|} \ :\ v \in T_\omega\Omega \setminus \{0\} \right\}.$$

Under Assumptions 3 and 4 :

$$\inf_{\bar{w}\in\mathfrak{S}} \lambda_1(\bar{w}) - \lambda_1(f,\mu) \le \int \ln\left[1 + \left(1 + C_{\phi,\Phi}(\omega)\right)\kappa_{\mathrm{ret}}\right] d\mu(\omega). \quad (3)$$

Two modes of learning

Prediction mode. Directly learn the dynamics at time k, i.e. find :

$$w_k : \mathbb{R}^{L+d} \to \mathbb{R}^d, \quad w_k\left(\Phi(x)\right) = \left(U^k \phi\right)(x) = \phi\left(f^k x\right).$$

Two modes of learning

Prediction mode. Directly learn the dynamics at time k, i.e. find :

$$w_k : \mathbb{R}^{L+d} \to \mathbb{R}^d, \quad w_k\left(\Phi(x)\right) = \left(U^k \phi\right)(x) = \phi\left(f^k x\right).$$

Reconstruction mode. Learn one step and iterate : Or one can iterate the step 1 learning k times :

$$w = w_1 : \mathbb{R}^{L+d} \to \mathbb{R}^d, \quad w\left(\Phi(x)\right) = \left(U\phi\right)(x) = \phi\left(fx\right).$$

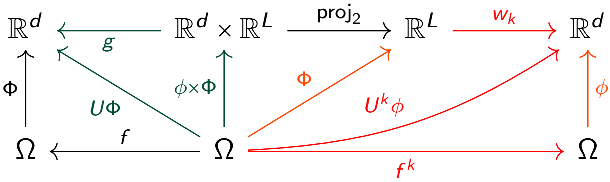
Two modes of learning

Prediction mode. Directly learn the dynamics at time k, i.e. find :

$$w_k : \mathbb{R}^{L+d} \to \mathbb{R}^d, \quad w_k\left(\Phi(x)\right) = \left(U^k \phi\right)(x) = \phi\left(f^k x\right).$$

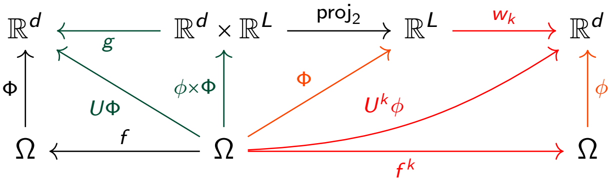
Reconstruction mode. Learn one step and iterate : Or one can iterate the step 1 learning k times :

$$w = w_1 : \mathbb{R}^{L+d} \to \mathbb{R}^d, \quad w\left(\Phi(x)\right) = \left(U\phi\right)(x) = \phi\left(fx\right).$$



Direct vs iterative

The predictive / direct mode, and the reconstrucive / iterative mode differ only in the choice of k in the following diagram.



However, as a learning problem, the growth of their errors are drastically different, governed by properties of the Koopman operator and the Lyapunov spectrum respectively.

Hypothesis space

The hypothesis space $\mathcal{H}$ will be a finite dimensional space, spanned by a basis $h_1, \ldots, h_m$. In that case

$$\mathcal{W} := \text{span}\left\{h_i \circ \Phi_l \;:\; 1 \le i \le m, 1 \le l \le L\right\} \tag{4}$$

WLOG we also assume that $\mathcal{W}$ contains the constant function $1_{\mathbb{R}^L}$. Define the *projection error* to be the quantity

$$\delta = \delta(\mathcal{H}) := \left\| (\text{Id} - \text{proj}_{\mathcal{W}})\, U\phi \right\|_{L^2(\mu)} . \tag{5}$$

If w_n is to be approximated by some $\hat{w}_n$ from $\mathcal{W}$.

$$\text{error}_{\text{direct}}(n) := \left\| U^n\phi - \hat{w}_n \circ \Phi \right\|_{L^2(\mu)} . \tag{6}$$

Theorem 2. Direct forecast

There is a subset $\mathbb{N}' \subseteq \mathbb{N}$ with density 1 such that :

Theorem 2. Direct forecast

There is a subset $\mathbb{N}' \subseteq \mathbb{N}$ with density 1 such that :

(i) For every $\epsilon > 0$, if $\mathcal{W}$ is large enough, then

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \|\phi - \text{proj}_{\mathcal{D}} \phi\|_{L^2(\mu)} + \epsilon.$$

Theorem 2. Direct forecast

There is a subset $\mathbb{N}' \subseteq \mathbb{N}$ with density 1 such that :

(i) For every $\epsilon > 0$, if $\mathcal{W}$ is large enough, then

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \left\| \phi - \text{proj}_{\mathcal{D}} \phi \right\|_{L^2(\mu)} + \epsilon.$$

(ii) If f is weakly mixing, then for every choice of $\mathcal{W}$

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \text{var}_\mu := \left\| \phi - \mu(\phi) \right\|_{L^2(\mu)}.$$

Theorem 2. Direct forecast

There is a subset $\mathbb{N}' \subseteq \mathbb{N}$ with density 1 such that :

(i) For every $\epsilon > 0$, if $\mathcal{W}$ is large enough, then

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \|\phi - \text{proj}_{\mathcal{D}} \phi\|_{L^2(\mu)} + \epsilon.$$

(ii) If f is weakly mixing, then for every choice of $\mathcal{W}$

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \text{var}_\mu := \|\phi - \mu(\phi)\|_{L^2(\mu)}.$$

(iii) If f is strongly mixing, the set $\mathbb{N}'$ can be taken to be the entire set $\mathbb{N}$.

Theorem 2. Direct forecast

There is a subset $\mathbb{N}' \subseteq \mathbb{N}$ with density 1 such that :

(i) For every $\epsilon > 0$, if $\mathcal{W}$ is large enough, then

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \|\phi - \text{proj}_{\mathcal{D}} \phi\|_{L^2(\mu)} + \epsilon.$$

(ii) If f is weakly mixing, then for every choice of $\mathcal{W}$

$$\lim_{n \in \mathbb{N}', n \to \infty} \text{error}_{\text{direct}}(n) = \text{var}_{\mu} := \|\phi - \mu(\phi)\|_{L^2(\mu)}.$$

(iii) If f is strongly mixing, the set $\mathbb{N}'$ can be taken to be the entire set $\mathbb{N}$.

(iv) If f has purely discrete spectrum, then for every $\epsilon > 0$, if the hypothesis space $\mathcal{W}$ is chosen large enough, then

$$\text{error}_{\text{direct}}(n) < \epsilon, \quad \forall n \in \mathbb{N}.$$

Theorem 3. Iterative forecast

$$\begin{aligned} \mathrm{error}_{\mathrm{iter}}(n,\omega) &:= \left\| U^n\phi(\omega) - \mathrm{proj}_1 \circ \hat{\mathcal{T}}^n \circ (\phi,\Phi)(\omega) \right\|_{\mathbb{R}^d}, \\ \mathrm{error}_{\mathrm{iter}}(n) &:= \left[\int_\Omega \mathrm{error}_{\mathrm{iter}}(n,\omega)^2 d\mu(\omega) \right]^{1/2}. \end{aligned} \tag{7}$$

Theorem 3. Iterative forecast

$$\begin{aligned} \mathrm{error}_{\mathrm{iter}}(n,\omega) &:= \left\| U^n \phi(\omega) - \mathrm{proj}_1 \circ \hat{\mathcal{T}}^n \circ (\phi, \Phi)(\omega) \right\|_{\mathbb{R}^d}, \\ \mathrm{error}_{\mathrm{iter}}(n) &:= \left[\int_\Omega \mathrm{error}_{\mathrm{iter}}(n,\omega)^2 d\mu(\omega) \right]^{1/2}. \end{aligned} \tag{7}$$

1. For every $\epsilon > 0$, there is a a constant $C^{(1)}_{\omega,\epsilon}$ such that

$$\mathrm{error}_{\mathrm{iter}}(n,\omega) = \| \Delta u_n(\omega) \|_{\mathbb{R}^L} = \delta C^{(1)}_{\omega,\epsilon} O\left(e^{n(\lambda_1 + \epsilon)} \right), \quad \text{as } n \to \infty, \tag{8}$$

Theorem 3. Iterative forecast

$$\begin{aligned} \text{error}_{\text{iter}}(n,\omega) &:= \left\| U^n \phi(\omega) - \text{proj}_1 \circ \hat{\mathcal{T}}^n \circ (\phi, \Phi)(\omega) \right\|_{\mathbb{R}^d}, \\ \text{error}_{\text{iter}}(n) &:= \left[\int_\Omega \text{error}_{\text{iter}}(n,\omega)^2 d\mu(\omega) \right]^{1/2}. \end{aligned} \tag{7}$$

1. For every $\epsilon > 0$, there is a a constant $C^{(1)}_{\omega,\epsilon}$ such that

$$\text{error}_{\text{iter}}(n,\omega) = \|\Delta u_n(\omega)\|_{\mathbb{R}^L} = \delta C^{(1)}_{\omega,\epsilon} O\left(e^{n(\lambda_1+\epsilon)}\right), \quad \text{as } n \to \infty, \tag{8}$$

2. If (Ω, μ, f) has L^2 Pesin sets , then for every $\epsilon > 0$,

$$\text{error}_{\text{iter}}(n) = \|\Delta u_n\|_{L^2(\mu)} = \delta C^{(2)}_{\epsilon} O\left(e^{n(\lambda_1+\epsilon)}\right), \quad \text{as } n \to \infty. \tag{9}$$

for a constant $C^{(2)}_\epsilon$ that depends only on ϵ.

An associated linear system

Consider matrix valued functions

$$\begin{aligned} W &: \Omega \to \mathbb{R}^{d\times L}, && W(\omega) := Dw \circ \Phi(\omega), \\ \hat{W} &: \Omega \to \mathbb{R}^{d\times L}, && \hat{W}(\omega) := D\hat{w}|_{\Phi(\omega)} = D\hat{w} \circ \Phi(\omega), \\ G^{(1)} &: \Omega \to \mathbb{R}^{L\times d}, && G^{(1)}(\omega) := \nabla_1 g|_{h(\omega)} = \nabla_1 g \circ h(\omega), \\ G^{(2)} &: \Omega \to \mathbb{R}^{L\times L}, && G^{(2)}(\omega) := \nabla_2 g|_{h(\omega)} \nabla_2 g \circ h(\omega), \end{aligned} \tag{10}$$

and their combination

$$M : \Omega \to \mathbb{R}^{(L+d)\times(L+d)}, \quad M(\omega) := \begin{bmatrix} 0^{d\times d} & W(\omega) \\ G^{(1)}(\omega) & G^{(2)}(\omega) \end{bmatrix}. \tag{11}$$

vector-valued functions

$$c : \Omega \to \mathbb{R}^L, \; c(\omega) := G^{(1)}(\omega) \left(U^{-1}\Delta\phi\right)(\omega).$$

Perturbed matrix cocycle

Thus associated to the dynamics are two functions :

$$M : \Omega \to \mathbb{R}^{(L+d)\times(L+d)}, \quad c : \Omega \to \mathbb{R}^{L}.$$

Perturbed matrix cocycle

Thus associated to the dynamics are two functions :

$$M : \Omega \to \mathbb{R}^{(L+d)\times(L+d)}, \quad c : \Omega \to \mathbb{R}^{L}.$$

Fix an $\omega \in \Omega$ and define

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = M(f^n\omega) \begin{bmatrix} a_n \\ b_n \end{bmatrix} + \begin{bmatrix} 0 \\ c(f^n\omega) \end{bmatrix}, \quad \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \vec{0} \qquad (12)$$

Perturbed matrix cocycle

Thus associated to the dynamics are two functions :

$$M : \Omega \to \mathbb{R}^{(L+d)\times(L+d)}, \quad c : \Omega \to \mathbb{R}^{L}.$$

Fix an $\omega \in \Omega$ and define

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = M\left(f^n\omega\right)\begin{bmatrix} a_n \\ b_n \end{bmatrix} + \begin{bmatrix} 0 \\ c\left(f^n\omega\right) \end{bmatrix}, \quad \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \vec{0} \qquad (12)$$

If one interprets the initial state ω as a random variable, then $M\left(f^n\omega\right)$, $c\left(f^n\omega\right)$ are random sequences of matrices and vectors, leading to the driven / random affine system (12).

Perturbed matrix cocycle

Thus associated to the dynamics are two functions :

$$M : \Omega \to \mathbb{R}^{(L+d)\times(L+d)}, \quad c : \Omega \to \mathbb{R}^{L}.$$

Fix an $\omega \in \Omega$ and define

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = M\left(f^n\omega\right)\begin{bmatrix} a_n \\ b_n \end{bmatrix} + \begin{bmatrix} 0 \\ c\left(f^n\omega\right) \end{bmatrix}, \quad \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \vec{0} \qquad (12)$$

If one interprets the initial state ω as a random variable, then $M\left(f^n\omega\right)$, $c\left(f^n\omega\right)$ are random sequences of matrices and vectors, leading to the driven / random affine system (12).
This dynamics involves products

$$M\left(f^n\omega\right) M\left(f^{n-1}\omega\right) \cdots M\left(\omega\right),$$

which brings in multiplicative ergodic theory.

Theorem 3 - revisited

The reconstruction gives the dynamics :

$$\begin{bmatrix} u_{n+1} \\ y_{n+1} \end{bmatrix} = \hat{\mathcal{T}} \begin{bmatrix} u_n \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{w}\left(y_n\right) \\ g\left(u_n, y_n\right) \end{bmatrix}. \tag{13}$$

Theorem 3 - revisited

The reconstruction gives the dynamics :

$$\begin{bmatrix} u_{n+1} \\ y_{n+1} \end{bmatrix} = \hat{\mathcal{T}} \begin{bmatrix} u_n \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{w}(y_n) \\ g(u_n, y_n) \end{bmatrix}. \quad (13)$$

We are interested in the deviations :

$$\begin{bmatrix} \Delta u_n \\ \Delta y_n \end{bmatrix} = \begin{bmatrix} U^{n-1} \pi U \phi \\ U^n \Phi \end{bmatrix} - \begin{bmatrix} u_n \\ y_n \end{bmatrix}, \quad \forall n \in \mathbb{N}_0. \quad (14)$$

Theorem 3 - revisited

The reconstruction gives the dynamics :

$$\begin{bmatrix} u_{n+1} \\ y_{n+1} \end{bmatrix} = \hat{\mathcal{T}} \begin{bmatrix} u_n \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{w}\left(y_n\right) \\ g\left(u_n, y_n\right) \end{bmatrix}. \tag{13}$$

We are interested in the deviations :

$$\begin{bmatrix} \Delta u_n \\ \Delta y_n \end{bmatrix} = \begin{bmatrix} U^{n-1} \pi U \phi \\ U^n \Phi \end{bmatrix} - \begin{bmatrix} u_n \\ y_n \end{bmatrix}, \quad \forall n \in \mathbb{N}_0. \tag{14}$$

Random affine dynamics from (12) :

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = M\left(f^n \omega\right) \begin{bmatrix} a_n \\ b_n \end{bmatrix} + \begin{bmatrix} 0 \\ c\left(f^n \omega\right) \end{bmatrix}, \quad \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \vec{0}$$

Theorem 3 - revisited

The reconstruction gives the dynamics :

$$\begin{bmatrix} u_{n+1} \\ y_{n+1} \end{bmatrix} = \hat{\mathcal{T}} \begin{bmatrix} u_n \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{w}\left(y_n\right) \\ g\left(u_n, y_n\right) \end{bmatrix}. \tag{13}$$

We are interested in the deviations :

$$\begin{bmatrix} \Delta u_n \\ \Delta y_n \end{bmatrix} = \begin{bmatrix} U^{n-1}\pi U\phi \\ U^n \Phi \end{bmatrix} - \begin{bmatrix} u_n \\ y_n \end{bmatrix}, \quad \forall n \in \mathbb{N}_0. \tag{14}$$

Random affine dynamics from (12) :

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = M\left(f^n\omega\right) \begin{bmatrix} a_n \\ b_n \end{bmatrix} + \begin{bmatrix} 0 \\ c\left(f^n\omega\right) \end{bmatrix}, \quad \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \vec{0}$$

The affine system (12) approximates the error growth of (13) :

$$\Delta u_n = a_n + O\left(a_{n-1}\right)^2, \quad \lim_{\delta \to 0} \frac{\left\|\Delta u_n\right\|}{\left\|a_n\right\|} = 1. \tag{15}$$

Conclusions

1. Error from direct forecasts decay at the rate of decay of correlations of the dynamical system, an ergodic property.

Conclusions

1. Error from direct forecasts decay at the rate of decay of correlations of the dynamical system, an ergodic property.
2. Error from direct forecasts decay at a rate determined by the Lyapunov exponents of the system.

Conclusions

1. Error from direct forecasts decay at the rate of decay of correlations of the dynamical system, an ergodic property.
2. Error from direct forecasts decay at a rate determined by the Lyapunov exponents of the system.
3. The stability of the reconstructed system depends on how the map w was learned in the ambient neighborhood of the image $\Phi(X)$ of the attractor.

Open problems

Devise learning techniques for w with either of these goals :

1. adapt the learning algorithm to inspect the unstable directions more closely than the stable directions.

Open problems

Devise learning techniques for w with either of these goals :

1. adapt the learning algorithm to inspect the unstable directions more closely than the stable directions.
2. preserve the stability of the attractor.

Open problems

Devise learning techniques for w with either of these goals :

1. adapt the learning algorithm to inspect the unstable directions more closely than the stable directions.
2. preserve the stability of the attractor.
3. Preserve the statistical property of the attractor.

Open problems

Devise learning techniques for w with either of these goals :

1. adapt the learning algorithm to inspect the unstable directions more closely than the stable directions.
2. preserve the stability of the attractor.
3. Preserve the statistical property of the attractor.
4. preserve the topology of the attractor.

References

1. T. Berry, S. Das, *Learning theory for dynamical systems*, **Arxiv:2208.05349** , 2022
2. L. Barreira, Ya. Pesin, *Lyapunov exponents and smooth ergodic theory*, 2002
3. MS. Raghunathan, *A proof of Oseledec's multiplicative ergodic theorem*, **Israel J. Math. 32**, 1979
4. D Giannakis, S. Das, J. Slawinska, *Reproducing kernel Hilbert space compactification of unitary evolution groups*, **Appl. Comput. Harm. Anal. 54** 2021.

Thank you !!