

# Learning Latent Dynamics for Control

Travaux dans le cadre de l'ANR Delicio<sup>1</sup>

Université Claude Bernard Lyon 1, CNRS, LAGEPP, Villeurbanne, France



Université Claude Bernard Lyon 1



---

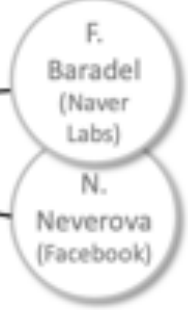
<sup>1</sup>M.Nadri; nadri@lagep.univ-lyon1.fr



Prediction of Physics

PhD of S. Janny

PhD of Q. Possamai



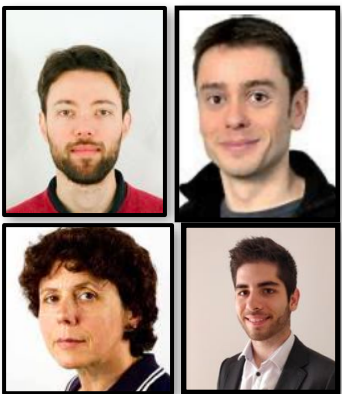
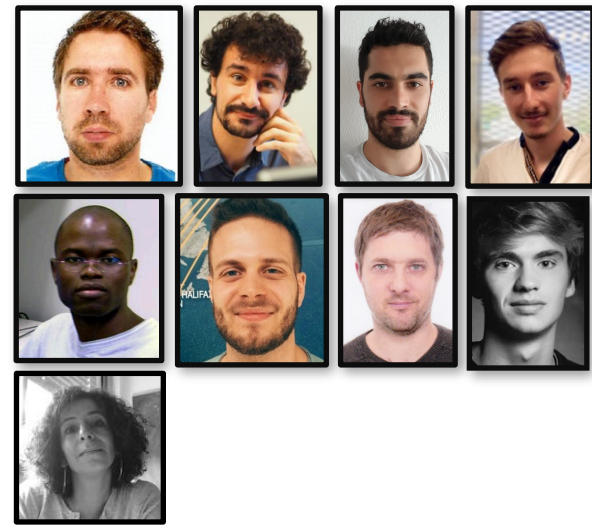
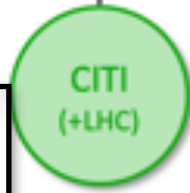
External collaborations

System identification in presence of turbulences



PhD of M. Marchand

PhD of S. Zoboli



Consider a nonlinear system of the general form

$$\begin{cases} x_{k+1} &= f(x_k, u_k) \\ y_k &= h(x_k, u_k) + w_k, \end{cases}$$

Results with theoretical guarantees:

- Model identification
- State estimation: Observer design.
- Stability and stabilisation analysis
- Optimal/robust Control design ...

Consider a nonlinear system of the general form

$$\begin{cases} x_{k+1} &= f(x_k, u_k) \\ y_k &= h(x_k, u_k) + w_k, \end{cases}$$

Results with theoretical guarantees:

- Model **identification**
- **State estimation**: Observer design.
- Stability and stabilisation **analysis**
- Optimal/robust **Control design** ...

## Classical challenges:

- How to obtain a representative model?
- How to obtain reliable measurements and for which price?
- How to obtain low computational complexity suitable for real-time ?

## Classical challenges:

- How to obtain a representative model?
- How to obtain reliable measurements and for which price?
- How to obtain low computational complexity suitable for real-time ?

## How does Control Theory proceed?

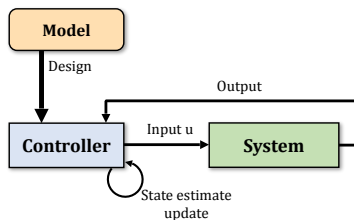
- Model reduction/simplification
- Linear (local) approximation
- Specific cases

## Classical challenges:

- How to obtain a representative model?
- How to obtain reliable measurements and for which price?
- How to obtain low computational complexity suitable for real-time ?

## How does Control Theory proceed?

- Model reduction/simplification
- Linear (local) approximation
- Specific cases



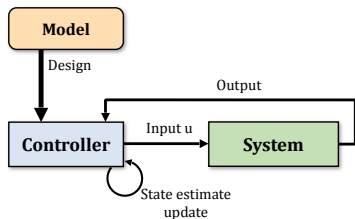
(a) The control theory viewpoint

## Classical challenges:

- How to obtain a representative model?
- How to obtain reliable measurements and for which price?
- How to obtain low computational complexity suitable for real-time ?

## How does Control Theory proceed?

- Model reduction/simplification
- Linear (local) approximation
- Specific cases



(a) The control theory viewpoint

- ✗ Local
- ✗ Can be conservative
- ✗ Domain dependent (requires an expert)

## Learning-based Methods:

- ✓ Arbitrary complex systems
- ✓ Arbitrarily large dimension
- ✓ Data driven (model-free): No domain expertise

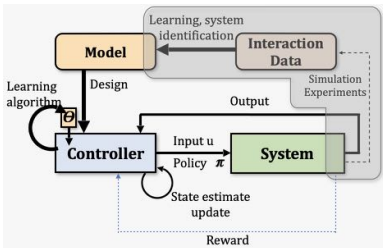
- ✗ "No" guarantees
- ✗ Black box
- ✗ Big data required
- ✗ Generalization issues.

## Learning-based Methods:

- ✓ Arbitrary complex systems
- ✓ Arbitrarily large dimension
- ✓ Data driven (model-free): No domain expertise

- ✗ "No" guarantees
- ✗ Black box
- ✗ Big data required
- ✗ Generalization issues.

## New Framework: Hybrid methods



- ✓ Combine physics and learning /gray box: better interpretability
- ✓ Complex and large dimension
- ✓ Needs less data
- ✓ Some theoretical guarantees

$x$ -coordinates in  $\mathbb{R}^{n_x}$

$$\begin{aligned}\dot{x} &= Fx \\ y &= Hx\end{aligned}$$

$$\hat{x} = ?$$

$z$ -coordinates in  $\mathbb{R}^{n_z}$

$$\dot{z} = Az + By$$

$A$  Hurwitz

$$\dot{\hat{z}} = A\hat{z} + By$$

We have

$$\frac{d}{dt}(\hat{z} - z) = A(\hat{z} - z) \implies \lim_{t \rightarrow \infty} \hat{z}(t) - z(t) = 0$$

and if  $T$  invertible,

$$\lim_{t \rightarrow \infty} \hat{x}(t) - x(t) = 0$$

<sup>2</sup>Observing the state of a linear system, D. Luenberger, 1964

Goal: find  $T$  invertible such that  $z = Tx$  verifies  $\dot{z} = Az + By$  with  $A$  Hurwitz

⇒  $T$  must verify the "Sylvester equation"

$$TF = AT + BH$$

- unique solution if  $F$  and  $A$  have no common eigenvalue
- invertible if  $(F, H)$  observable and  $(A, B)$  controllable

⇒ Luenberger observer :

$$\dot{\hat{z}} = A\hat{z} + By \quad , \quad \hat{x} = T^{-1}\hat{z}$$

$$\iff$$

$$\dot{\hat{x}} = F\hat{x} + T^{-1}B(y - H\hat{x})$$

Extension to nonlinear systems: **KKL Observer**<sup>3</sup>

---

<sup>3</sup>[Kazantzis, C. Kravaris], *Nonlinear observer design using Lyapunov's auxiliary theorem*, SCL, 1998

Goal: find  $T$  invertible such that  $z = Tx$  verifies  $\dot{z} = Az + By$  with  $A$  Hurwitz

$\Rightarrow T$  must verify the "Sylvester equation"

$$TF = AT + BH$$

- unique solution if  $F$  and  $A$  have no common eigenvalue
- invertible if  $(F, H)$  observable and  $(A, B)$  controllable

$\Rightarrow$  Luenberger observer :

$$\dot{\hat{z}} = A\hat{z} + By, \quad \hat{x} = T^{-1}\hat{z}$$



$$\dot{\hat{x}} = F\hat{x} + T^{-1}B(y - H\hat{x})$$

Extension to nonlinear systems: **KKL Observer**<sup>3</sup>

---

<sup>3</sup>[Kazantzis, C. Kravaris], *Nonlinear observer design using Lyapunov's auxiliary theorem*, SCL, 1998

Goal: find  $T$  invertible such that  $z = Tx$  verifies  $\dot{z} = Az + By$  with  $A$  Hurwitz

$\Rightarrow T$  must verify the "Sylvester equation"

$$TF = AT + BH$$

- unique solution if  $F$  and  $A$  have no common eigenvalue
- invertible if  $(F, H)$  observable and  $(A, B)$  controllable

$\Rightarrow$  Luenberger observer :

$$\dot{\hat{z}} = A\hat{z} + By, \quad \hat{x} = T^{-1}\hat{z}$$



$$\dot{\hat{x}} = F\hat{x} + T^{-1}B(y - H\hat{x})$$

Extension to nonlinear systems: KKL Observer<sup>3</sup>

---

<sup>3</sup>[Kazantzis, C. Kravaris], *Nonlinear observer design using Lyapunov's auxiliary theorem*, SCL, 1998

Goal: find  $T$  invertible such that  $z = Tx$  verifies  $\dot{z} = Az + By$  with  $A$  Hurwitz

$\Rightarrow T$  must verify the "Sylvester equation"

$$TF = AT + BH$$

- unique solution if  $F$  and  $A$  have no common eigenvalue
- invertible if  $(F, H)$  observable and  $(A, B)$  controllable

$\Rightarrow$  Luenberger observer :

$$\dot{\hat{z}} = A\hat{z} + By \quad , \quad \hat{x} = T^{-1}\hat{z}$$



$$\dot{\hat{x}} = F\hat{x} + T^{-1}B(y - H\hat{x})$$

Extension to nonlinear systems: KKL Observer<sup>3</sup>

<sup>3</sup>[Kazantzis, C. Kravaris], *Nonlinear observer design using Lyapunov's auxiliary theorem*, SCL, 1998

Goal: find  $T$  invertible such that  $z = Tx$  verifies  $\dot{z} = Az + By$  with  $A$  Hurwitz

$\Rightarrow T$  must verify the "Sylvester equation"

$$TF = AT + BH$$

- unique solution if  $F$  and  $A$  have no common eigenvalue
- invertible if  $(F, H)$  observable and  $(A, B)$  controllable

$\Rightarrow$  Luenberger observer :

$$\dot{\hat{z}} = A\hat{z} + By \quad , \quad \hat{x} = T^{-1}\hat{z}$$

$$\iff$$

$$\dot{\hat{x}} = F\hat{x} + T^{-1}B(y - H\hat{x})$$

Extension to nonlinear systems: **KKL Observer**<sup>3</sup>

<sup>3</sup>[Kazantzis, C. Kravaris], *Nonlinear observer design using Lyapunov's auxiliary theorem*, SCL, 1998

- 1 KKL observer based deep learning
  - Problem statement
  - KKL observer design
  - Numerical construction of the mapping
- 2 KKL Based Deep Learning
  - Algorithm extension
  - Illustration examples
- 3 Output predictor: Deep-KKL based approach
  - Prediction problem
  - Output prediction based data
  - Simulation results
- 4 Conclusion and Future Works

We consider the non-autonomous discrete-time system

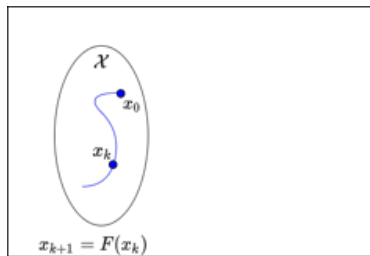
$$\begin{cases} x_{k+1} = F(x_k, u_k), & x \in \mathbb{R}^{d_x}, u \in \mathbb{R}^{d_u} \\ y_k = h(x_k), & y \in \mathbb{R}^{d_y} \end{cases} \quad (1)$$

- Luenberger observers (KKL): find an injective and continuous map  $T$  which transforms the system into a Hurwitz form.

We consider the non-autonomous discrete-time system

$$\begin{cases} x_{k+1} = F(x_k, u_k), & x \in \mathbb{R}^{d_x}, u \in \mathbb{R}^{d_u} \\ y_k = h(x_k), & y \in \mathbb{R}^{d_y} \end{cases} \quad (1)$$

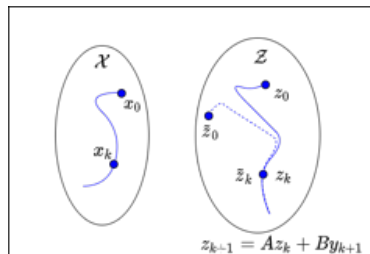
- **Luenberger observers (KKL)**: find an injective and continuous map  $T$  which transforms the system into a Hurwitz form.



We consider the non-autonomous discrete-time system

$$\begin{cases} x_{k+1} = F(x_k, u_k), & x \in \mathbb{R}^{d_x}, u \in \mathbb{R}^{d_u} \\ y_k = h(x_k), & y \in \mathbb{R}^{d_y} \end{cases} \quad (1)$$

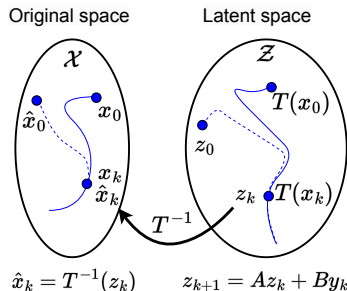
- **Luenberger observers (KKL)**: find an injective and continuous map  $T$  which transforms the system into a Hurwitz form.



We consider the non-autonomous discrete-time system

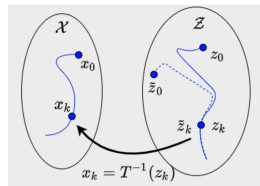
$$\begin{cases} x_{k+1} = F(x_k, u_k), & x \in \mathbb{R}^{d_x}, u \in \mathbb{R}^{d_u} \\ y_k = h(x_k), & y \in \mathbb{R}^{d_y} \end{cases} \quad (1)$$

- **Luenberger observers (KKL)**: find an injective and continuous map  $T$  which transforms the system into a Hurwitz form.



## Problems:

- $T$  is difficult to compute and we need its pseudo inverse  $T^{-1}$  to design the observer in the original space.
- It is assumed that the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .



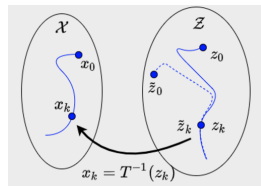
## Our Approach<sup>4</sup>: Construct the map $T$ and its pseudo inverse.

- Algorithm based on ensemble learning technique
- An extension of the mapping  $\mathcal{T}$  outside  $T(\mathcal{X})$  in the transient phase
- A high dimension systems application

<sup>4</sup>[Peralez, Nadri CDC, 2021.]

## Problems:

- $T$  is difficult to compute and we need its pseudo inverse  $T^{-1}$  to design the observer in the original space.
- It is assumed that the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .



## Our Approach<sup>4</sup>: Construct the map $T$ and its pseudo inverse.

- Algorithm based on ensemble learning technique
- An extension of the mapping  $\mathcal{T}$  outside  $T(\mathcal{X})$  in the transient phase
- A high dimension systems application

<sup>4</sup>[Peralez, Nadri CDC, 2021.]

## Assuming:

- There exists a compact set  $\mathcal{X}$  such that for any solution  $x$  to equation 1,  $x_k \in \mathcal{X}$  for all  $k \geq 0$ .
- For all  $u^0$  of interest,  $F(., u^0)$  is invertible and  $F^{-1}$ ,  $h$  are  $C^1$  and globally Lipschitz.
- $\forall x_1 \neq x_2 \implies \exists i \text{ s.t. } h(F^{-i}(x_1)) \neq h(F^{-i}(x_2))$

## Theorem (Brivadis et al, 2019)

For any pair of matrices  $A \in \mathcal{R}^{d_z \times d_z}$  and  $B \in \mathcal{R}^{d_z \times d_y}$  there exists a continuous mapping  $T : \mathcal{X} \rightarrow \mathbb{R}^{d_z}$  satisfying

$$T(F(x, u^0)) = AT(x) + Bh(x) \quad \forall x \in \mathcal{X}. \quad (2)$$

If  $d_z \geq d_y(d_x + 1)$ ,  $(A, B)$  is controllable and  $A$  is Schur stable, then there exists  $T$  injective, and the (left inverse) mapping  $T^{-1}$  exists.

A KKL observer is given by:

$$\begin{cases} z_{k+1} = Az_k + By_k \\ \hat{x}_k = T^{-1}(z_k), \end{cases} \quad (3)$$

Properties:

- Thanks to the contraction properties, we have the convergence in the latent space.
- Since  $T$  is injective, a convergent estimation in the original state space is guaranteed.

Questions:

- How to "generically" construct the mappings  $T$  and its inverse ?
- How to improve the robustness to noise?
- How to handle the transient phase?

If  $d_z \geq d_y(d_x + 1)$ ,  $(A, B)$  is controllable and  $A$  is Schur stable, then there exists  $T$  injective, and the (left inverse) mapping  $T^{-1}$  exists.

A KKL observer is given by:

$$\begin{cases} z_{k+1} = Az_k + By_k \\ \hat{x}_k = T^{-1}(z_k), \end{cases} \quad (3)$$

Properties:

- Thanks to the contraction properties, we have the convergence in the latent space.
- Since  $T$  is injective, a convergent estimation in the original state space is guaranteed.

Questions:

- How to "generically" construct the mappings  $T$  and its inverse ?
- How to improve the robustness to noise?
- How to handle the transient phase?

If  $d_z \geq d_y(d_x + 1)$ ,  $(A, B)$  is controllable and  $A$  is Schur stable, then there exists  $T$  injective, and the (left inverse) mapping  $T^{-1}$  exists.

A KKL observer is given by:

$$\begin{cases} z_{k+1} = Az_k + By_k \\ \hat{x}_k = T^{-1}(z_k), \end{cases} \quad (3)$$

Properties:

- Thanks to the contraction properties, we have the convergence in the latent space.
- Since  $T$  is injective, a convergent estimation in the original state space is guaranteed.

Questions:

- **How to "generically" construct the mappings  $T$  and its inverse ?**
- **How to improve the robustness to noise?**
- **How to handle the transient phase?**

Recall the considered system

$$\begin{cases} x_{k+1} &= F(x_k, u_k) \\ y_k &= h(x_k) \end{cases}$$

Lets denote

$$\mathcal{T}(T(x)) = x \quad \forall x \in \mathcal{X}.$$

A KKL observer is given by<sup>5</sup>

$$\begin{cases} \hat{z}_{k+1} &= A\hat{z}_k + By_k + \Psi(\hat{z}, u_k) \\ \hat{x}_k &= \mathcal{T}(\hat{z}_k) \end{cases} \quad (4)$$

with

$$\Psi(\hat{z}, u_k) := T(F(\mathcal{T}(\hat{z}_k), u_k)) - T(F(\mathcal{T}(\hat{z}_k), u^0)).$$

By choosing the eigenvalues of  $A$  fast enough the estimation error exponentially converges.

---

<sup>5</sup>[Peralez et al, CDC 2021]

Recall the considered system

$$\begin{cases} x_{k+1} &= F(x_k, u_k) \\ y_k &= h(x_k) \end{cases}$$

Lets denote

$$\mathcal{T}(T(x)) = x \quad \forall x \in \mathcal{X}.$$

A KKL observer is given by<sup>5</sup>

$$\begin{cases} \hat{z}_{k+1} &= A\hat{z}_k + By_k + \Psi(\hat{z}, u_k) \\ \hat{x}_k &= \mathcal{T}(\hat{z}_k) \end{cases} \quad (4)$$

with

$$\Psi(\hat{z}, u_k) := T(F(\mathcal{T}(\hat{z}_k), u_k)) - T(F(\mathcal{T}(\hat{z}_k), u^0)).$$

By choosing the eigenvalues of  $A$  fast enough the estimation error exponentially converges.

---

<sup>5</sup>[Peralez et al, CDC 2021]

Recall the considered system

$$\begin{cases} x_{k+1} &= F(x_k, u_k) \\ y_k &= h(x_k) \end{cases}$$

Lets denote

$$\mathcal{T}(T(x)) = x \quad \forall x \in \mathcal{X}.$$

A KKL observer is given by<sup>5</sup>

$$\begin{cases} \hat{z}_{k+1} &= A\hat{z}_k + By_k + \Psi(\hat{z}, u_k) \\ \hat{x}_k &= \mathcal{T}(\hat{z}_k) \end{cases} \quad (4)$$

with

$$\Psi(\hat{z}, u_k) := T(F(\mathcal{T}(\hat{z}_k), u_k)) - T(F(\mathcal{T}(\hat{z}_k), u^0)).$$

By choosing the eigenvalues of  $A$  fast enough the estimation error exponentially converges.

---

<sup>5</sup>[Peralez et al, CDC 2021]

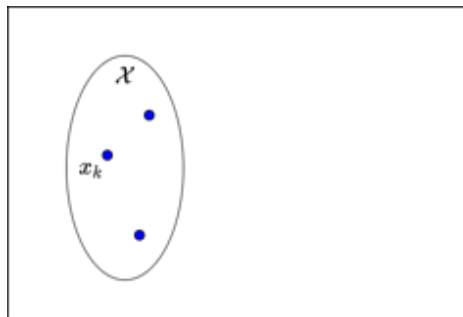
- **Sample  $x_k$** : from a uniform distribution on  $\mathcal{X}$ .
- **Sample  $x_{k+1}, y_{k+1}$** : from model simulation fixed control input  $u_0$ .
- **Learn the mapping  $T$** : minimize the following loss:

$$\mathcal{L}_k = |T(x_{k+1}) - (AT(x_k) + By_k)|.$$

- **Find the mapping  $\mathcal{T}$** : minimize the following loss:

$$\mathcal{L}_{\mathcal{T}} = |x_k - \mathcal{T}(T(x_k))|.$$

- Assumption: the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .
- Problem: if  $z$  is outside  $T(\mathcal{X})$ ,  $\mathcal{T}$  is not defined outside  $T(\mathcal{X})$ .



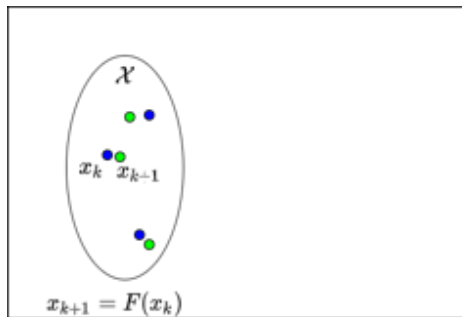
- Sample  $x_k$ : from a uniform distribution on  $\mathcal{X}$ .
- Sample  $x_{k+1}, y_{k+1}$ : from model simulation fixed control input  $u_0$ .
- Learn the mapping  $T$ : minimize the following loss:

$$\mathcal{L}_k = |T(x_{k+1}) - (AT(x_k) + By_k)|.$$

- Find the mapping  $\mathcal{T}$ : minimize the following loss:

$$\mathcal{L}_{\mathcal{T}} = |x_k - \mathcal{T}(T(x_k))|.$$

- Assumption: the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .
- Problem: if  $z$  is outside  $T(\mathcal{X})$ ,  $\mathcal{T}$  is not defined outside  $T(\mathcal{X})$ .



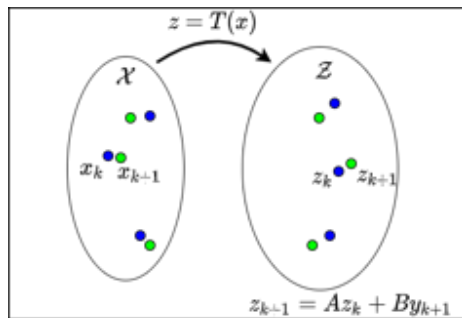
- Sample  $x_k$ : from a uniform distribution on  $\mathcal{X}$ .
- Sample  $x_{k+1}, y_{k+1}$ : from model simulation fixed control input  $u_0$ .
- Learn the mapping  $T$ : minimize the following loss:

$$\mathcal{L}_k = |T(x_{k+1}) - (AT(x_k) + By_k)|.$$

- Find the mapping  $\mathcal{T}$ : minimize the following loss:

$$\mathcal{L}_{\mathcal{T}} = |x_k - \mathcal{T}(T(x_k))|.$$

- Assumption: the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .
- Problem: if  $z$  is outside  $T(\mathcal{X})$ ,  $\mathcal{T}$  is not defined outside  $T(\mathcal{X})$ .

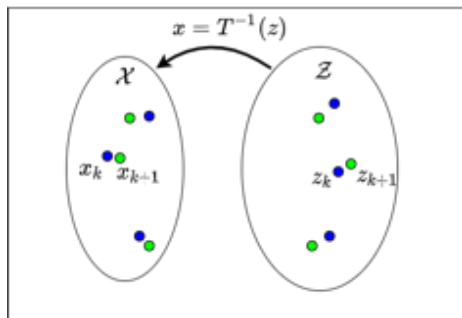


- Sample  $x_k$ : from a uniform distribution on  $\mathcal{X}$ .
- Sample  $x_{k+1}, y_{k+1}$ : from model simulation fixed control input  $u_0$ .
- Learn the mapping  $T$ : minimize the following loss:

$$\mathcal{L}_k = |T(x_{k+1}) - (AT(x_k) + By_k)|.$$

- Find the mapping  $\mathcal{T}$ : minimize the following loss:

$$\mathcal{L}_{\mathcal{T}} = |x_k - \mathcal{T}(T(x_k))|.$$



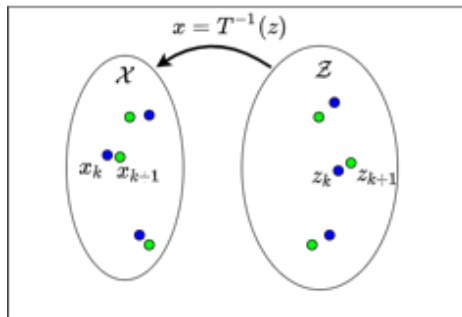
- Assumption: the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .
- Problem: if  $z$  is outside  $T(\mathcal{X})$ ,  $\mathcal{T}$  is not defined outside  $T(\mathcal{X})$ .

- Sample  $x_k$ : from a uniform distribution on  $\mathcal{X}$ .
- Sample  $x_{k+1}, y_{k+1}$ : from model simulation fixed control input  $u_0$ .
- Learn the mapping  $T$ : minimize the following loss:

$$\mathcal{L}_k = |T(x_{k+1}) - (AT(x_k) + By_k)|.$$

- Find the mapping  $\mathcal{T}$ : minimize the following loss:

$$\mathcal{L}_{\mathcal{T}} = |x_k - \mathcal{T}(T(x_k))|.$$



- Assumption: the observer state in the latent space  $\hat{z}$  remains in  $T(\mathcal{X})$ .
- Problem: if  $z$  is outside  $T(\mathcal{X})$ ,  $\mathcal{T}$  is not defined outside  $T(\mathcal{X})$ .

- We consider the mapping extension:

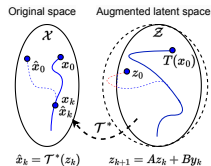
$$\tilde{T}(z) = \operatorname{argmin}_{x \in \mathcal{X}} |z - T(x)| \quad \forall z \in \mathbb{R}^{d_z}. \quad (5)$$

- The new observer for the non-autonomous system is then given by:

$$\begin{cases} \hat{z}_{k+1} &= A\hat{z}_k + By_k + \tilde{\Psi}(\hat{z}_k, u_k) \\ \hat{x}_k &= \tilde{T}(\hat{z}_k), \end{cases} \quad (6)$$

where now  $\tilde{\Psi}$  is defined as:

$$\tilde{\Psi}(\hat{z}_k, u_k) := T(F(\tilde{T}(\hat{z}_k), u_k)) - T(F(\tilde{T}(\hat{z}_k), u^0)).$$



## Proposition

Lets suppose there exists a positive constant  $c_\Psi$  satisfying <sup>a</sup>

$$|\tilde{\Psi}(z^1, u) - \tilde{\Psi}(z^2, u)| \leq c_\Psi |z^1 - z^2|$$

for all  $u \in \mathcal{U}$  and for all  $z^1, z^2 \in T(\mathcal{X})$ , such that the matrix  $A + 2c_\Psi I$  is Schur stable. Then any solution to (1), equation 5, equation 6 verifies  $\lim_{k \rightarrow +\infty} |x_k - \hat{x}_k| = 0$ .

<sup>a</sup>[Peralez et al CDC, 2022]

- We consider the mapping extension:

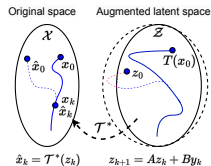
$$\tilde{T}(z) = \operatorname{argmin}_{x \in \mathcal{X}} |z - T(x)| \quad \forall z \in \mathbb{R}^{d_z}. \quad (5)$$

- The new observer for the non-autonomous system is then given by:

$$\begin{cases} \hat{z}_{k+1} &= A\hat{z}_k + By_k + \tilde{\Psi}(\hat{z}_k, u_k) \\ \hat{x}_k &= \tilde{T}(\hat{z}_k), \end{cases} \quad (6)$$

where now  $\tilde{\Psi}$  is defined as:

$$\tilde{\Psi}(\hat{z}_k, u_k) := T(F(\tilde{T}(\hat{z}_k), u_k)) - T(F(\tilde{T}(\hat{z}_k), u^0)).$$



## Proposition

Lets suppose there exists a positive constant  $c_\Psi$  satisfying <sup>a</sup>

$$|\tilde{\Psi}(z^1, u) - \tilde{\Psi}(z^2, u)| \leq c_\Psi |z^1 - z^2|$$

for all  $u \in \mathcal{U}$  and for all  $z^1, z^2 \in T(\mathcal{X})$ , such that the matrix  $A + 2c_\Psi I$  is Schur stable. Then any solution to (1), equation 5, equation 6 verifies  $\lim_{k \rightarrow +\infty} |x_k - \hat{x}_k| = 0$ .

<sup>a</sup>[Peralez et al CDC, 2022]

- Compute an (over) approximation of a set  $\mathcal{Z}$ : estimate the image of  $T(\mathcal{X})$  using the loss:

$$\mathcal{L}_{\tilde{\mathcal{T}}} = |z_k - T(\tilde{\mathcal{T}}(z_k))| + w d(\tilde{\mathcal{T}}(z_k), \mathcal{X})^2,$$

- $d(x_k, \mathcal{X})$  is the distance between  $\tilde{\mathcal{T}}(z_k)$  and the compact set  $\mathcal{X}$ .
- State estimation:

$$\hat{x} = \begin{cases} \tilde{\mathcal{T}}(\hat{z}), & \text{if } |\hat{z} - T(\mathcal{T}(\hat{z}))| > \epsilon_{\mathcal{T}} \\ \mathcal{T}(\hat{z}), & \text{otherwise.} \end{cases}$$

- Extension to Multi-KKL approach<sup>6</sup>: Improve the transient response

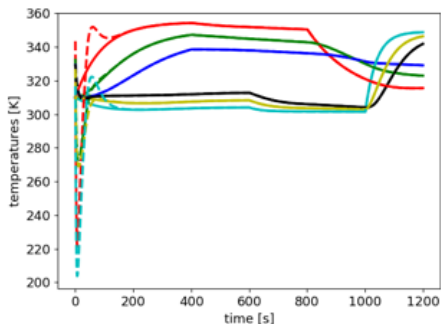
---

<sup>6</sup>[Peralez, Nadri, CDC, 2022]

- **A deep auto-encoder (MLP)**: encodes  $x$  into a latent space  $z=T(x)$  and decodes the representation to recover the coordinates  $x$ .
- **Algorithm 1**: a base model training, where the mappings  $T$ ,  $\mathcal{T}$  and  $\tilde{\mathcal{T}}$  are trained sequentially.
- The **learning rate** used in the gradient descent is gradually adapted.
- **Algorithm 2**: Combining predictions from multiple models s.t. Each call of Algorithm 1 is done with an optimized version of the mapping parameters, allowing to reduce the number of iterations using Snapshot ensemble method.

- **The PDE counter-current heat exchanger model:**

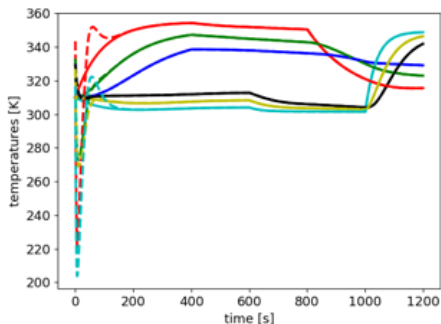
$$\begin{aligned}
 T_k &= -v_h T_s - \kappa_h (T - \bar{T}) \\
 \bar{T}_k &= v_c \bar{T}_s - \kappa_c (\bar{T} - T) \\
 T(t, 0) &= T_{in}, \quad \bar{T}(t, 1) = \bar{T}_{in}, \\
 T(0, s) &= T_0(s), \quad \bar{T}(0, s) = \bar{T}_0(s),
 \end{aligned}$$



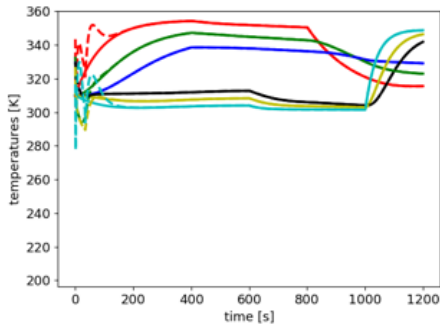
Baseline method.

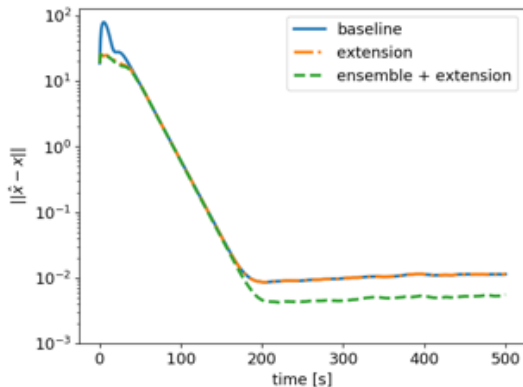
- The PDE counter-current heat exchanger model:

$$\begin{aligned}
 T_k &= -v_h T_s - \kappa_h (T - \bar{T}) \\
 \bar{T}_k &= v_c \bar{T}_s - \kappa_c (\bar{T} - T) \\
 T(t, 0) &= T_{in}, \quad \bar{T}(t, 1) = \bar{T}_{in}, \\
 T(0, s) &= T_0(s), \quad \bar{T}(0, s) = \bar{T}_0(s),
 \end{aligned}$$



Baseline method.

Benefits of the extension of  $\tilde{T}$  during the transient phase.



Benefits of the extension of MKKI.

- **The PDE Korteweg-de Vries model:**

$$\begin{cases} w_t + w_x + w_{xxx} + ww_x = 0 \\ w(t, 0) = u(t), w_x(t, L) = 0, w_{xx}(t, L) = 0 \\ w(0, x) = w_0(x) \\ y(t) = w(t, L). \end{cases}$$

- The spatial discretization with  $d_x = 35$  points on  $[0, 2\pi]$  and a time step of 0.01 s.

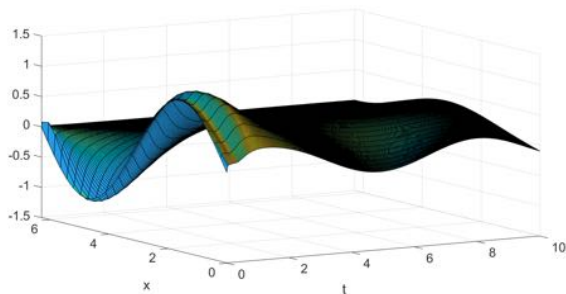


Figure: KdV equation - An example of trajectory.

# Nonlinear observer for the Korteweg-de Vries model

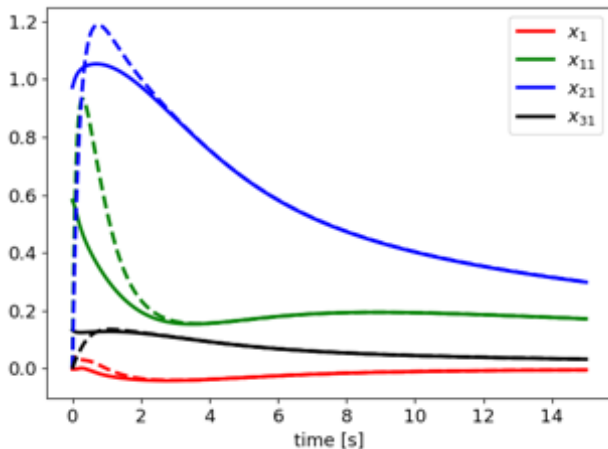


Figure: KdV equation - an estimation result with 5 base estimators and a switching threshold fixed to  $\epsilon_{\mathcal{T}} = 10^{-4}$

Output predictor: Deep-KKL approach<sup>7</sup>

---

<sup>7</sup>[Janny et al. , CDC 2021]

Consider a process/natural phenomena that generates a (real) time function

$$\mathbb{Y} = \{t \in \mathcal{R} \mapsto y(t) \in \mathcal{R}\}.$$

**Prediction problem:** *Given a current time  $t$  can we infer the future value of an experiment  $y$  in  $\mathbb{Y}$  given that we know  $y(s)$ , for  $s$  in  $[0, t]$  ?*

Standard problem in various fields of engineering :

- Meteorology
- Stocks values
- Engineering
- In control for MPC

We may try to solve this problem for a subset  $\mathcal{Y} \subseteq \mathbb{Y}$ .

Consider a process/natural phenomena that generates a (real) time function

$$\mathbb{Y} = \{t \in \mathcal{R} \mapsto y(t) \in \mathcal{R}\}.$$

**Prediction problem:** *Given a current time  $t$  can we infer the future value of an experiment  $y$  in  $\mathbb{Y}$  given that we know  $y(s)$ , for  $s$  in  $[0, t]$  ?*

Standard problem in various fields of engineering :

- Meteorology
- Stocks values
- Engineering
- In control for MPC

We may try to solve this problem for a subset  $\mathbf{Y} \subseteq \mathbb{Y}$ .

We are looking for a predictor in the form

$$y_p(t + p) = \psi(z(t)) , \dot{z} = F(z, y(t)) , z \in \mathcal{R}^m$$

- $p$  is the prediction horizon
- $z$  is the latent state which aggregates information.

In M.L. this is solved with RNN or GRU.

**Example:** a RNN of depth 1 is

$$y_p(t + p) = \psi(z(t)) , z_{k+1} = \sigma(W_0 z_k + W_1 y_k + b)$$

- with  $\sigma$  a diagonal sigmoid function
- $\psi$ ,  $W_0$ ,  $W_1$  and  $b$  are learnt on a set of data  $Y_D \subset \mathbb{Y}$ .

We consider the deterministic dynamical system

$$\dot{x} = f(x) , \quad y = h(x) ,$$

with  $f : \mathcal{R}^n \mapsto \mathcal{R}^n$ ,  $h : \mathcal{R}^n \mapsto \mathcal{R}$ , and  $X(x_0, t)$  is the unique and complete solution.

- The experiment is fully described by the initial condition  $x_0$  in  $\mathcal{R}^n$
- The output is explained by a **FINITE** dimensional deterministic dynamical system

$$\mathbb{Y} = \{y : \mathcal{R}^+ \mapsto \mathcal{R}, \exists x_0 , y(s) = h(X(x_0, s))\}.$$

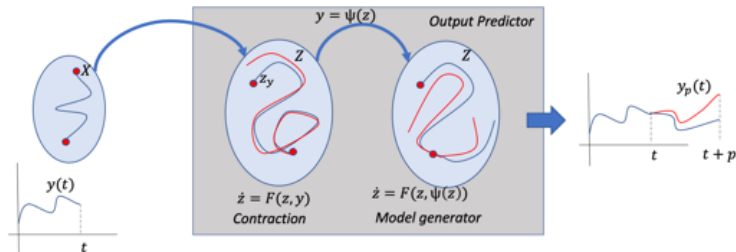
## Output Predictor for $\mathbf{Y} \subset \mathbb{Y}$

is defined as a couple  $(F, \psi)$  such as

- $\dot{z} = F(z, y)$  is a uniform exponential contraction with parameter  $(k, \lambda)$  ;
- the couple  $(f, \psi)$  with  $f(z) = F(z, \psi(z))$  is a generating model for  $\mathbf{Y}$ .

the output predictor is given as

$$y_p(t+p) = \psi(\mathbf{Z}(z(t), p)) , \dot{z}(t) = F(z(t), y(t)) , z(0) = 0$$



## Proposition

Assume there exist  $F$  and  $\psi$ , such that  $(F, \psi)$  defines an output predictor for  $\mathbf{Y}$  with:

$$\left| \frac{\partial f}{\partial z}(z) \right| \leq L_1, \quad \left| \frac{\partial \psi}{\partial z}(z) \right| \leq L_2,$$

with  $f(z) = F(z, \psi(z))$ , then for all experiments  $y \in \mathbf{Y}$ ,

$$y_p(t+p) = \psi(\mathbf{Z}(z(t), p)), \quad \dot{z}(t) = F(z(t), y(t)), \quad z(0) = 0$$

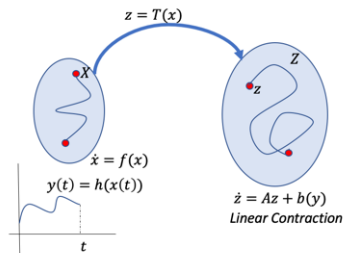
satisfies

$$|y_p(t+p) - y(t+p)| \leq kL_2 e^{-\lambda t + L_1 p} |z_0^y|.$$

- As the prediction horizon increases, the prediction error grows as well.
- For each fixed prediction horizon, the upper-bound exponentially goes to zero for increasing  $t$ .

**Question:** When and how can we design an output predictor ?

Step 1 for KKL Predictor: Sending a measured dynamical system into a linear contraction

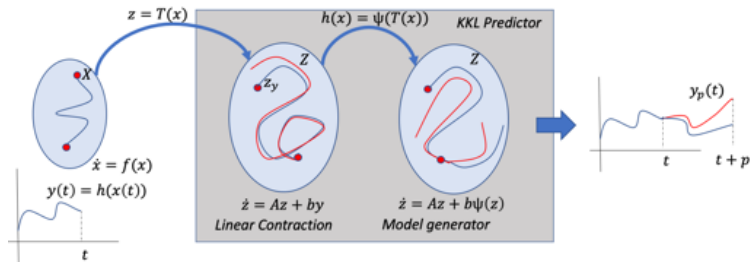


# A possible solution via KKL

Step 2 for KKL Predictor: If there exists  $\psi$  such that  $h(x) = \psi(T(x))$  then

$$\dot{z} = Az + b\psi(z) , y = \psi(z)$$

generates the output



Question: When does  $\psi$  exist?

## Theorem

With  $m = 2n + 1$ , for all (diagonal) stable matrices  $A$  and vector  $b$  with  $(A, b)$  controllable there exists a continuous mapping  $\psi : R^m \mapsto R$  such that,  $(F, \psi)$  defines an output predictor for  $Y_O$ .

- No observability assumption is needed
- The proof relies on [Andrieu et Praly, SIAM 2006], [Marconi et al. SIAM 2007].

Given  $f$  and  $h$ ,

- 1 Select  $m$  sufficiently large
- 2 Select  $(A, b)$  controllable with  $A$  stable
- 3 Compute the solution to the PDE :

$$L_f T(x) = AT(x) + bh(x)$$

- 4 Compute a (Lipschitz) function  $\psi$  such that

$$\psi(T(x)) = h(x)$$

The predictor is

$$y_p(t + p) = \psi(\mathbf{Z}(z(t), p)) , \dot{z}(t) = Az(t) + by(t)$$

Question : What can be done if  $(f, h)$  is unknown ?

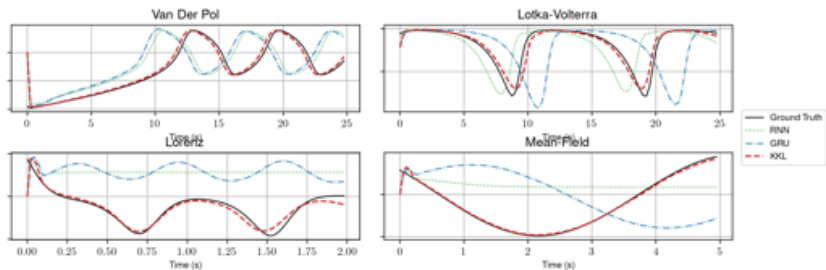
$$\dot{x} = f(x), \quad y = h(x),$$

- $(f, h)$  is unknown
- We know a set of experiments  $Y_D \subset \mathbb{Y}$
- We model  $\psi_\theta$ , as a MLP  $\theta \in \Theta \subset \mathbb{R}^q$ ,  $q \gg 1$
- Pick  $b = (1, \dots, 1)$
- The predictor is defined as

$$y_p(t+p) = \psi_\theta(z_t), \quad z_{t+1} = Az_t + by_t,$$

where

$$(\theta, A) = \arg \min_{\theta, A \text{ stable}} \sum_{y \in Y_D} \sum_{s=0}^{t+p} \|y(s) - \psi_\theta(z(s))\|^2.$$



- New **hybrid methodologies** for modeling, estimation and control.
- **Robustness analysis and guarantees** available in spite of data-driven components.
- Address **systems of high dimension**: two MIMO nonlinear PDE models where the explicit solution is practically impossible.

### Future work:

- Optimizing the latent dynamics.
- Relax assumptions in non-autonomous case.
- Incremental training and adaptivity of RNN with Input-to-State Stability guarantees
- Realization of large-scale finite-dimensional systems from infinite dimensions that preserve some desired properties.
- Which NN-based control architecture to provide guarantees for emulated control strategies?

- New **hybrid methodologies** for modeling, estimation and control.
- **Robustness analysis and guarantees** available in spite of data-driven components.
- Address **systems of high dimension**: two MIMO nonlinear PDE models where the explicit solution is practically impossible.

### Future work:

- Optimizing the latent dynamics.
- Relax assumptions in non-autonomous case.
- Incremental training and adaptivity of RNN with Input-to-State Stability guarantees
- Realization of large-scale finite-dimensional systems from infinite dimensions that preserve some desired properties.
- Which NN-based control architecture to provide guarantees for emulated control strategies?