Output predictor: An approach based on control theory and data-driven techniques

V. Andrieu

CNRS - Université Claude Bernard de Lyon - LAGEPP - DYCOP

10/06/2021

Work in collaboration with S. Janny, C. Wolf and M. Nadri

Delicio ANR Project

The output predictor problem

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

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

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 $Y \subseteq \mathbb{Y}$

The output predictor problem

Various methods depending on the problem

We are looking for a predictor in the form

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

- p is the prediction horizon
- z is the latent state which agregates informations

In AI this is solved with Recurrent Neural Network (RNN) or Gated Recurrent Unit (GRU).

Example : a RNN of depth 1 is :

$$y_p(t+p) = \psi(z(t)) \ , \ \dot{z}(t) = \sigma(W_0 z(t) + W_1 y(t) + b)$$

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

The deterministic assumption

Claude Bernard

Il y a un déterminisme absolu dans les conditions d'existence des phénomènes naturels

The output is explained by a FINITE dimensional deterministic dynamical system

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

where, $X(x_0, t)$ is the unique and complete solution of

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

with $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, $h : \mathbb{R}^n \mapsto \mathbb{R}$

- the *world* dimension is $n >> 1$
- The experiment is fully described by the initial condition x_0 in $\mathbb{R}^n$

Outlines of the presentation

In this talk,

1. A general method based on
   1. Contraction
   2. Generating model
2. Existence of a KKL predictor
3. Data-based approach
4. Conclusion

Contraction

Consider

$$\dot{z} = F(z, y),$$

where z in $\mathbb{R}^m$ and y in $\mathbb{R}$.

Definition : Uniform Exponential Contraction

There exists (k, λ) such that for each solutions $z_a(\cdot)$, $z_b(\cdot)$

$$|z_a(t) - z_b(t)| \leq ke^{-\lambda t} |z_a(0) - z_b(0)| \ , \ \forall (y, t)$$

- A contraction forgets its initial condition.
- All solutions converge to the same one which depends only on y

A generating model

Let $Y \subseteq \mathbb{Y}$

Consider

$$\dot{z} = \mathfrak{f}(z) \ , \ y = \psi(z),$$

where z in $\mathbb{R}^m$ and y in $\mathbb{R}$ with solution $\mathcal{Z}(z_0, t)$.

Definition : Generating Model for Y

is defined as a couple $(\mathfrak{f}, \psi)$ such that for all y in Y there exists z_0^y in $\mathbb{R}^m$ such that $y(t) = \psi(\mathcal{Z}(z_0^y, t))$.

- (f, h) is a generating model.
- If we know z_0^y, we can predict the output by integrating the generating model
- This is an internal model

Output Predictor

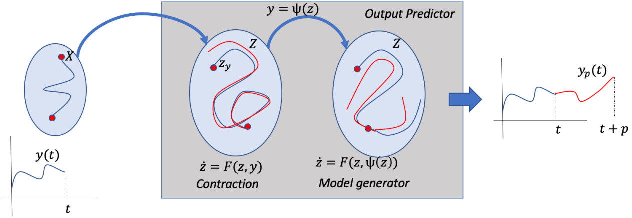
Output Predictor for $Y \subset \mathbb{Y}$

is defined as a couple (F, ψ) such as

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

the output predictor is given as

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



Output Predictor

Proposition

Assume there exist F and ψ, such that (F, ψ) defines an output predictor for Y with:

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

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

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

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 ?

Outlines of the presentation

In this talk,

1. A general method based on
   1. Contraction
   2. Generating model
2. Existence of a KKL predictor
3. Data-based approach
4. Conclusion

A possible solution via KKL

The framework

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

where, $X(x_0, t)$ is the unique and complete solution of

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

Assume that there exists $\mathcal{O} \subset \mathbb{R}^n$ bounded and invariant, ie. for all x_0 in $\mathcal{O}$:

$$X(x_0, t) \in \mathcal{O} \ , \forall t \ , \in \mathbb{R}.$$

Question : Can we find a predictor for

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

KKL Observer approach gives a solution

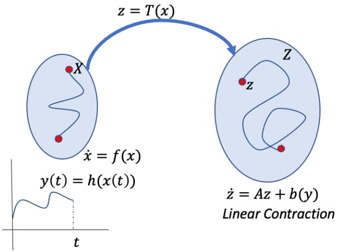
KKL Observer

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

If $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is solution to

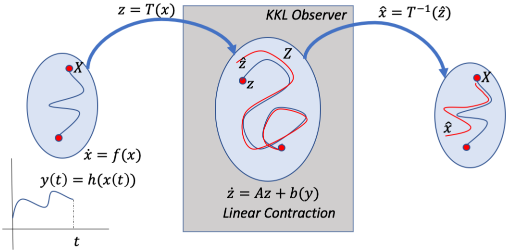
$$L_f T(x) = AT(x) + b(h(x))$$

then $z_y(t) = T(X(x,t))$ is solution to $\dot{z}(t) = Az(t) + b(y(t))$



KKL Observer

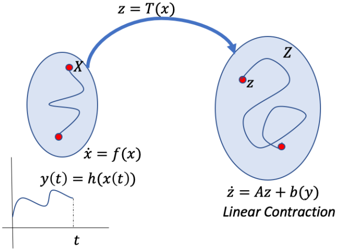
Step 2 for KKL Observer : Copy the dynamics and apply a left inverse



- The observer obtained is just a copy of the dynamics
- A KKL observer exists if the system is observable and m is sufficiently large

A possible solution via KKL

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

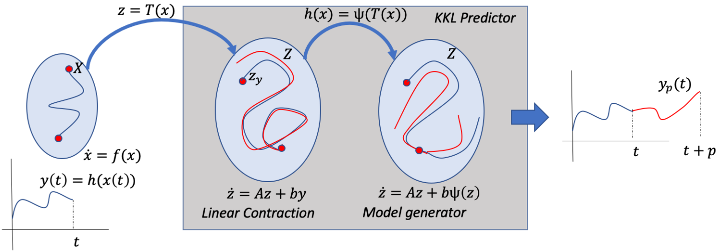


A possible solution via KKL

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

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

generates the output



Question : When does ψ exists ?

Existence of KKL Predictor

Theorem

With $m = 2n + 1$, for almost all (diagonal) stable matrix A and vector b with (A, b) controllable there exists a continuous mapping $\psi : \mathbb{R}^m \mapsto \mathbb{R}$ such that, (F, ψ) defines an output predictor for $Y_{\mathcal{O}}$.

- No observability assumption is needed
- The proof relies on Andrieu-Praly-2006, Marconi-Praly-Isidori-2007, Brivadis-Serres-Andrieu-Bernard-2021

A possible solution via KKL

- It is always possible to *explain* an output with a dynamical system in the form

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

  with A stable.

- Comparison with Koopman approach
  - The latent space created by the Koopman operator contains information about the full state.
  - The latent space is of finite dimension.

Question : What about the Lipschitz property of ψ ?

ψ Lipschitz

Proposition

Assume

- *Backward Distinguishability:* $\forall\ x_1 \neq x_2$ in $\mathcal{O}^2$ there exists $t \leq 0$ such that $h(X(x_1, t)) \neq h(X(x_2, t))$
- *Backward Infinitesimal Distinguishability:* $\forall\ (x, v)$ in $\mathcal{O} \times \mathbb{R}^n$ such that $v \neq 0$, there exists $t \leq 0$ such that

$$\frac{\partial h(X(x,t))}{\partial x} v \neq 0$$

then for almost all (diagonal) stable A, vector b with (A, b) controllable, there exists a ψ, C^1 and $L_2 > 0$ such that

1. Deep-KKL defines an output predictor for $Y_{\mathcal{O}}$;
2. The function ψ is globally Lipschitz

$$\left|\frac{\partial \psi}{\partial z}(z)\right| \leq L_2\ ,\ \forall z \in \mathbb{R}^m;$$

3. then

$$|y_p(t+p) - y(t+p)| \leq k L_2 e^{-\lambda t + L_1 p} |z_0^y|$$

with $L_1 = \|A\| + \|b\| L_2$.

ψ Lipschitz

To obtain the Lipschitz property, we need an observability assumption.

Existence of coordinates such that the system takes a triangular form

$$\left\{\begin{array}{l}\dot{x}_1 = f_1(x_1)\\ \dot{x}_2 = f_2(x_1, x_2)\end{array}\right. \quad , \quad y = h(x_1) \ ,$$

for which the couple (f_1, h) satisfies the observability assumptions of the proposition is sufficient.

Some geometric conditions are given in Isidori-Praly-Marconi-2010

Construction of the predictor

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 ψ such that

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

The predictor is

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

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

Outlines of the presentation

In this talk,

1. A general method based on
   1. Contraction
   2. Generating model
2. Existence of a KKL predictor
3. Data-based approach
4. Conclusion

Data based approach

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

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

where

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

Comparison with existing algorithm

- RNN:

$$z_{t+1} = \tanh(W_1 z_t + W_2 y_t + b)$$

- GRU:

$$\begin{aligned} r_{t+1} &= \sigma(W_{r1} y_t + W_{r2} z_t + b_r) \\ x_{t+1} &= \sigma(W_{x1} y_t + W_{x2} z_t + b_x) \\ n_{t+1} &= \tanh(W_{n1} y_t + r_{t+1} * (W_{n2} z_t + b_{n2}) + b_{n1}) \\ z_{t+1} &= (1 - x_{t+1}) * n_{t+1} + x_{t+1} * z_t, \end{aligned}$$

where $\sigma(x) = (1 + e^{-x})^{-1}$

- KKL:

$$z_{t+1} = A z_t + b y_t \ , \ A \text{ stable}$$

with

$$y_p(t+p) = \psi_\theta(z_t)$$

Experiments

Van der Pol Oscillator

$$\begin{cases} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= (1 - x_1^2)x_2 - x_1 \end{cases}$$

Lorenz Attractor

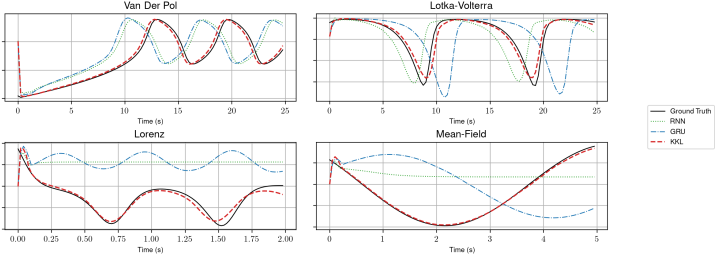
$$\begin{cases} \dot{x}_1 &= 10(x_2 - x_1) \\ \dot{x}_2 &= 24x_1 - x_2 - x_1x_3 \\ \dot{x}_3 &= x_1x_2 - \frac{8}{3}x_3 \end{cases}$$

Lotka-Volterra Equations

$$\begin{cases} \dot{x}_1 &= x_1(\frac{2}{3} - \frac{3}{4}x_2) \\ \dot{x}_2 &= x_2(x_1 - 1) \end{cases}$$

Mean-Field

$$\begin{cases} \dot{x}_1 &= 0.1x_1 - x_2 - 0.1x_1x_3 \\ \dot{x}_2 &= x_1 + 0.1x_2 - 0.1x_2x_3 \\ \dot{x}_3 &= -10(x_3 - x_1^2 - x_2^2) \end{cases}$$



- The $t = 5$ first time step were used in the closed loop behavior of each models
- The open-loop predicts the $p = 95$ following measurements

Global Performances

Mean Squared Error (MSE) on prediction

$$\mathcal{L}_{\text{MSE}} = \frac{1}{Np} \sum_{y \in Y_T} \sum_{s=t}^{t+p} (y(s) - \hat{y}(s))^2$$

where Y_T is the test set of trajectories, of cardinality N.

	RNN	GRU	KKL
Van Der Pol	0.0057	0.0343	0.**0013**
Lotka-Volterra	**0.0885**	0.1780	0.1064
Lorenz	0.0441	0.0480	**0.0262**
Mean-Field	0.2254	0.2044	**0.0012**

Conclusion

- Creating a dynamical system which is a contraction and a generating model is very promising approach for prediction
- When considering linear contraction it can be shown that there exists a predictor
- For low dimensional dynamical systems, numerical results are very good
- RNN dynamics may define a contraction

$$\dot{z}(t) = \sigma(W_0 z(t) + W_1 y(t) + b)$$

  if σ monotonic, this may define a contraction. Does there exists ψ ?
- What about systems with input ?
- What would be the best contraction ?