Data-driven Prediction and Control Through the Lens of System Identification

Andrea Iannelli

Automatic Control Lab, ETH Zürich, Switzerland

Trends on Dissipativity in Systems and Control - Brig - May 23, 2022

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

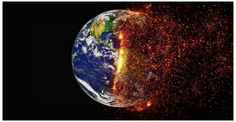
INSTITUT FÜR AUTOMATIK

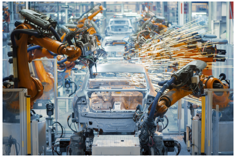
Swiss National Science Foundation

Paradigm shift in Engineering

- From component-oriented to system-level design (smart cities, intelligent transportation, industrial automation..)

- From first-principles modelling to data-driven



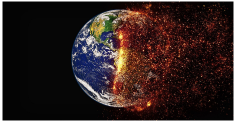


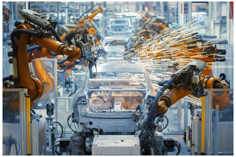
Paradigm shift in Engineering

- From component-oriented to system-level design (smart cities, intelligent transportation, industrial automation..)

- From first-principles modelling to data-driven

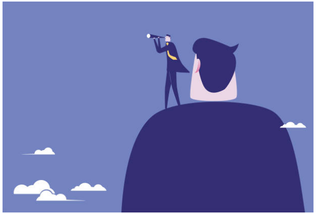
- **Methods** and **tools** must adapt, but not the **foundational principles**





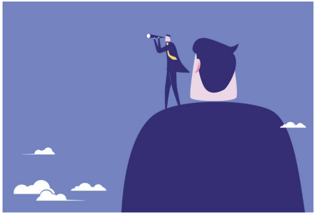
Research philosophy

- Standing on the shoulders of **giants** (optimal, robust, adaptive control, system identification, dynamical systems theory)..



Research philosophy

- Standing on the shoulders of **giants** (optimal, robust, adaptive control, system identification, dynamical systems theory)..
- ..seeking new peaks (statistical learning, optimization, information theory)



(New) Direct data-driven methods through the lens of a(n old) giant

Methods for simulating and controlling dynamical systems **directly** from data

$\rightarrow$ no knowledge of an explicit model is required

- Behavioral system theory (noise-free data)

(New) Direct data-driven methods through the lens of a(n old) giant

Methods for simulating and controlling dynamical systems **directly** from data

$\rightarrow$ no knowledge of an explicit model is required

- Behavioral system theory (noise-free data)
- Recent works towards a stochastic counterpart [1-2]

[1] Pan Ou Faulwasser, *"On a Stochastic Fundamental Lemma and Its Use for Data-Driven MPC"* 2021

[2] Furieri Guo Martin Ferrari-Trecate, *"Near-Optimal Design of Safe Output Feedback Controllers from Noisy Data"* 2021.

(New) Direct data-driven methods through the lens of a(n old) giant

Methods for simulating and controlling dynamical systems **directly** from data

$\rightarrow$ no knowledge of an explicit model is required

- Behavioral system theory (noise-free data)
- Recent works towards a stochastic counterpart [1-2]
- Today: System identification-based statistical reasoning
- Framework for simulation, filtering and control from noisy data

Joint work with

Mingzhou Yin

Roy S. Smith

[1] Pan Ou Faulwasser, *"On a Stochastic Fundamental Lemma and Its Use for Data-Driven MPC"* 2021

[2] Furieri Guo Martin Ferrari-Trecate, *"Near-Optimal Design of Safe Output Feedback Controllers from Noisy Data"* 2021.

Contents of the talk

1. A Bayesian perspective on the Fundamental Lemma
2. Confidence regions for noisy predictions
3. Informative data for simulation problems

Behavioural system theory

Representation-free perspective on dynamical systems as sets of trajectories [Willems, 1987]

- The *Fundamental Lemma* [Willems, 2005] arguably a driving force behind recently proposed subspace-type data-driven analysis and control methods
- Non-parametric representation of an LTI system by the image of a Hankel data matrix

$$z_{[i,j]} \to \mathcal{H}_L(z_{[i,j]}) := \begin{bmatrix} z_i & z_{i+1} & \cdots & z_{N+i-L} \\ z_{i+1} & z_{i+2} & \cdots & z_{N+i-L+1} \\ \vdots & \vdots & & \vdots \\ z_{i+L-1} & z_{i+L} & \cdots & z_{N+i-1} \end{bmatrix}$$

  $z_{[i,j]}$ is: **persistently exciting** (PE) of order L if $\mathcal{H}_L(z_{[i,j]})$ has full row rank
- Can we use raw data to span the I/O behavior of the system?

Spanning the behavior?

Minimal linear time-invariant system

$$
\begin{aligned}
x_{t+1} &= Ax_t + Bu_t \\
y_t &= Cx_t + Du_t
\end{aligned}
$$

$x \in \mathbb{R}^{n_x}$ is the state, $u \in \mathbb{R}^{n_u}$ is the input, and $y \in \mathbb{R}^{n_y}$ is the output (the *lag* l is the smallest integer i such that the ext. observability matrix $\mathcal{O}_i$ has rank n_x).

Direct prediction/simulation problem

Given: an input-output **data trajectory** $(u_{\mathrm{d}\,[0,N-1]}, y_{\mathrm{d}\,[0,N-1]})$.

Goal: for any *input* **simulation trajectory** $u_{\mathrm{s}\,[0,L_s-1]}$, and any initial condition, find the (unique) *output* **simulation trajectory** $y_{\mathrm{s}\,[0,L_s-1]}$.

Fundamental lemma

[Willems, 2005]

Controllable LTI system with McMillan degree n_x, I/O trajectory $(u_{\mathrm{d}\ [0,N-1]}, y_{\mathrm{d}\ [0,N-1]})$. *If* $u_{\mathrm{d}\ [0,N-1]}$ is PE of order $L + n_x$, then $(u_{\mathrm{t}\ [0,L-1]}, y_{\mathrm{t}\ [0,L-1]})$ is an I/O trajectory **iff** there exists g s.t.

$$\begin{bmatrix} u_{\mathrm{t}\ [0,L-1]} \\ y_{\mathrm{t}\ [0,L-1]} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u_{\mathrm{d}\ [0,N-1]}) \\ \mathcal{H}_L(y_{\mathrm{d}\ [0,N-1]}) \end{bmatrix} g$$

Fundamental lemma & direct simulation

[Willems, 2005]

Controllable LTI system with McMillan degree n_x, I/O trajectory $(u_{\mathrm{d}\,[0,N-1]}, y_{\mathrm{d}\,[0,N-1]})$. *If* $u_{\mathrm{d}\,[0,N-1]}$ is PE of order $L + n_x$, then $(u_{\mathrm{t}\,[0,L-1]}, y_{\mathrm{t}\,[0,L-1]})$ is an I/O trajectory **iff** there exists g s.t.

$$\begin{bmatrix} u_{\mathrm{t}\,[0,L-1]} \\ y_{\mathrm{t}\,[0,L-1]} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u_{\mathrm{d}\,[0,N-1]}) \\ \mathcal{H}_L(y_{\mathrm{d}\,[0,N-1]}) \end{bmatrix} g$$

Given $L_0 \geq l$ (define $L_s = L - L_0$), $y_{\mathrm{s}\,[0,L_s-1]}$ ($\mathrm{y_s}$) is the unique output trajectory with past trajectory $(u_{\mathrm{ini}\,[0,L_0-1]}, y_{\mathrm{ini}\,[0,L_0-1]})$ ($\mathrm{u_{ini}}, \mathrm{y_{ini}}$) and input trajectory $u_{\mathrm{s}\,[0,L_s-1]}$ ($\mathrm{u_s}$), **iff** there exists g s.t.

$$\left[\begin{array}{c} \mathrm{u_{ini}} \\ \mathrm{u_s} \\ \hline \mathrm{y_{ini}} \\ \mathrm{y_s} \end{array}\right] = \left[\begin{array}{c} U \\ \hline Y \end{array}\right] g = \left[\begin{array}{c} U_p \\ U_f \\ \hline Y_p \\ Y_f \end{array}\right] g = \left[\begin{array}{c} \mathcal{H}_L(u_{\mathrm{d}\,[0,N-1]}) \\ \hline \mathcal{H}_L(y_{\mathrm{d}\,[0,N-1]}) \end{array}\right] g$$

Noise-free case abstraction

The trajectory can be partitioned into a known and unknown part

$$\begin{bmatrix} \hat{z} \\ \hline z \end{bmatrix} = \begin{bmatrix} u \\ \hline y \end{bmatrix}, \quad Z = \begin{bmatrix} U \\ Y \end{bmatrix}, \quad \left(\text{e.g. } \hat{z} = \begin{bmatrix} \color{blue}{u_{ini}} \\ \color{blue}{u_s} \\ \color{blue}{y_{ini}} \end{bmatrix}, z = \begin{bmatrix} \color{red}{y_s} \end{bmatrix} \right)$$

$\hat{z} = Z_1 g \rightarrow g^*(\hat{z}, Z_1)$, from: known and *data*, to g

$z = Z_2 g^* = Z_2 g^*(\hat{z}, Z_1)$ from: g and *data*, to unknown

Noise-free case abstraction

The trajectory can be partitioned into a known and unknown part

$$\begin{bmatrix} \hat{z} \\ \hline z \end{bmatrix} = \begin{bmatrix} u \\ \hline y \end{bmatrix}, \quad Z = \begin{bmatrix} U \\ Y \end{bmatrix}, \quad \left(\text{e.g. } \hat{z} = \begin{bmatrix} u_{ini} \\ u_s \\ y_{ini} \end{bmatrix}, z = \begin{bmatrix} y_s \end{bmatrix} \right)$$

$$\hat{z} = Z_1 g \rightarrow g^*(\hat{z}, Z_1), \qquad \text{from: known and } data\text{, to } g$$

$$z = Z_2 g^* = Z_2 g^*(\hat{z}, Z_1) \qquad \text{from: } g \text{ and } data\text{, to unknown}$$

$$z = \mathcal{F}_Z(\hat{z}), \quad (y_s = \mathcal{F}_Z(u_{ini}, u_s, y_{ini}) \ \textit{input-output mapping})$$

Noisy case *regression* viewpoint

$$\begin{aligned}
\tilde{y}_i^{\mathrm{d}} &= y_i^{\mathrm{d}} + w_i^{\mathrm{d}}, & w_i^{\mathrm{d}} &\sim \mathcal{N}(0, \sigma_y^2 I_p)\\
\tilde{u}_i^{\mathrm{d}} &= u_i^{\mathrm{d}} + v_i^{\mathrm{d}}, & v_i^{\mathrm{d}} &\sim \mathcal{N}(0, \sigma_u^2 I_m)\\
\tilde{y}_i^{\mathrm{ini}} &= y_i^{\mathrm{ini}} + w_i^{\mathrm{ini}}, & w_i^{\mathrm{ini}} &\sim \mathcal{N}(0, \sigma_{p,y}^2 I_p)\\
\tilde{u}_i^{\mathrm{ini}} &= u_i^{\mathrm{ini}} + v_i^{\mathrm{ini}}, & v_i^{\mathrm{ini}} &\sim \mathcal{N}(0, \sigma_{p,u}^2 I_m)
\end{aligned}$$

Noisy case *regression* viewpoint

$$
\begin{aligned}
\tilde{y}_i^{\mathrm{d}} &= y_i^{\mathrm{d}} + w_i^{\mathrm{d}}, & w_i^{\mathrm{d}} &\sim \mathcal{N}(0, \sigma_y^2 I_p) \\
\tilde{u}_i^{\mathrm{d}} &= u_i^{\mathrm{d}} + v_i^{\mathrm{d}}, & v_i^{\mathrm{d}} &\sim \mathcal{N}(0, \sigma_u^2 I_m) \\
\tilde{y}_i^{\mathrm{ini}} &= y_i^{\mathrm{ini}} + w_i^{\mathrm{ini}}, & w_i^{\mathrm{ini}} &\sim \mathcal{N}(0, \sigma_{p,y}^2 I_p) \\
\tilde{u}_i^{\mathrm{ini}} &= u_i^{\mathrm{ini}} + v_i^{\mathrm{ini}}, & v_i^{\mathrm{ini}} &\sim \mathcal{N}(0, \sigma_{p,u}^2 I_m)
\end{aligned}
$$

$$
\boxed{\underbrace{\hat{z}}_{\textit{generalized}\text{ measurement}} = \underbrace{z}_{\text{unknown trajectory}} + \underbrace{w}_{\textit{generalized}\text{ disturbance}}}
$$

Noisy case *regression* viewpoint

$$\begin{aligned}
\tilde{y}_i^{\mathrm{d}} &= y_i^{\mathrm{d}} + w_i^{\mathrm{d}}, & w_i^{\mathrm{d}} &\sim \mathcal{N}(0, \sigma_y^2 I_p) \\
\tilde{u}_i^{\mathrm{d}} &= u_i^{\mathrm{d}} + v_i^{\mathrm{d}}, & v_i^{\mathrm{d}} &\sim \mathcal{N}(0, \sigma_u^2 I_m) \\
\tilde{y}_i^{\mathrm{ini}} &= y_i^{\mathrm{ini}} + w_i^{\mathrm{ini}}, & w_i^{\mathrm{ini}} &\sim \mathcal{N}(0, \sigma_{p,y}^2 I_p) \\
\tilde{u}_i^{\mathrm{ini}} &= u_i^{\mathrm{ini}} + v_i^{\mathrm{ini}}, & v_i^{\mathrm{ini}} &\sim \mathcal{N}(0, \sigma_{p,u}^2 I_m)
\end{aligned}$$

$$\boxed{\underbrace{\hat{z}}_{\textit{generalized}\text{ measurement}} = \underbrace{z}_{\text{unknown trajectory}} + \underbrace{w}_{\textit{generalized}\text{ disturbance}}}$$

$z \sim \mathcal{N}(Zg, \Sigma_z(g))$	"Fundamental Lemma"-prior
$w \sim \mathcal{N}(0, \Sigma_w)$	Disturbance model
$\hat{z}\|z \sim \mathcal{N}(z, \Sigma_w)$	Likelihood

$$\Sigma_z(g) = \begin{bmatrix} \Sigma_{U_p}(g) & 0 & 0 & 0 \\ 0 & \Sigma_{U_f}(g) & 0 & 0 \\ 0 & 0 & \Sigma_{Y_p}(g) & 0 \\ 0 & 0 & 0 & \Sigma_{Y_f}(g) \end{bmatrix}$$

The Signal Matrix Model (SMM)

The maximum-a-posteriori (MAP) estimate of the trajectory is

$$
\begin{aligned}
\mathrm{z}^*(g) &= \arg\max_{\mathrm{z}} \underbrace{p(\mathrm{z}|\hat{\mathrm{z}})}_{\propto\, p(\hat{\mathrm{z}}|\mathrm{z})p(\mathrm{z})} \\
&= \Sigma_z(g)\left(\Sigma_z(g)+\Sigma_w\right)^{-1}\hat{\mathrm{z}} + \Sigma_w\left(\Sigma_z(g)+\Sigma_w\right)^{-1} Zg
\end{aligned}
\tag{MAP}
$$

Trajectory identification from noisy data

The Signal Matrix Model (SMM)

The maximum-a-posteriori (MAP) estimate of the trajectory is

$$z^*(g) = \arg\max_{z} \underbrace{p(z|\hat{z})}_{\propto\, p(\hat{z}|z)p(z)} = \Sigma_z(g)\left(\Sigma_z(g) + \Sigma_w\right)^{-1}\hat{z} + \Sigma_w\left(\Sigma_z(g) + \Sigma_w\right)^{-1} Zg \tag{MAP}$$

where g is obtained through maximum marginal-likelihood (ML)

$$g^* \in \arg\min_{g} \underbrace{-p(\hat{z}|g)}_{\log\det\left(\Sigma_z(g) + \Sigma_w\right) + \left(\hat{z} - Zg\right)^\top \left(\Sigma_z(g) + \Sigma_w\right)^{-1}\left(\hat{z} - Zg\right)} \tag{ML}$$

The Signal Matrix Model (SMM)

The maximum-a-posteriori (MAP) estimate of the trajectory is

$$\begin{aligned} z^*(g) &= \arg\max_{z} \underbrace{p(z|\hat{z})}_{\propto\, p(\hat{z}|z)p(z)} \\ &= \Sigma_z(g)\left(\Sigma_z(g)+\Sigma_w\right)^{-1}\hat{z} + \Sigma_w\left(\Sigma_z(g)+\Sigma_w\right)^{-1} Zg \end{aligned} \tag{MAP}$$

where g is obtained through maximum marginal-likelihood (ML)

$$g^* \in \arg\min_{g} \underbrace{-p(\hat{z}|g)}_{\log\det\left(\Sigma_z(g)+\Sigma_w\right) + (\hat{z}-Zg)^\top\left(\Sigma_z(g)+\Sigma_w\right)^{-1}(\hat{z}-Zg)} \tag{ML}$$

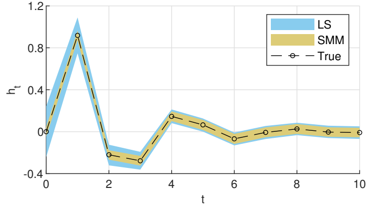
ML: from <u>known</u> and *data*, to g $\rightarrow$ MAP: from g and *data*, to <u>unknown</u>.

Example 1: Direct data-driven simulation

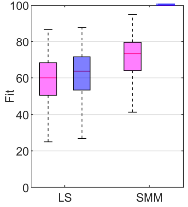
$$\hat{z} = \begin{bmatrix} \mathrm{u}_{\mathrm{ini}} \\ \mathrm{u}_{\mathrm{s}} \\ \mathrm{y}_{\mathrm{ini}} \\ 0 \end{bmatrix}, \; z = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathrm{y}_{\mathrm{s}} \end{bmatrix}, \; \Sigma_w = \begin{bmatrix} \Sigma_{u_i} & 0 & 0 & 0 \\ 0 & \Sigma_{u_s} & 0 & 0 \\ 0 & 0 & \Sigma_{y_i} & 0 \\ 0 & 0 & 0 & \Sigma_{y_s} \end{bmatrix}, \; \Sigma_{u_i} = \Sigma_{u_s} = \Sigma_{y_i} = 0, \; \Sigma_{y_s} = \infty \rightarrow \quad \mathrm{y}_{\mathrm{s}} = Y_f g^*$$

$$\hat{z} = \begin{bmatrix} u_{\text{ini}} \\ u_{\text{s}} \\ y_{\text{ini}} \\ 0 \end{bmatrix}, \; z = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y_{\text{s}} \end{bmatrix}, \; \Sigma_w = \begin{bmatrix} \Sigma_{u_i} & 0 & 0 & 0 \\ 0 & \Sigma_{u_s} & 0 & 0 \\ 0 & 0 & \Sigma_{y_i} & 0 \\ 0 & 0 & 0 & \Sigma_{y_s} \end{bmatrix}, \; \Sigma_{u_i} = \Sigma_{u_s} = \Sigma_{y_i} = 0, \; \Sigma_{y_s} = \infty \rightarrow \quad y_{\text{s}} = Y_f g^*$$

Truncated Infinite Impulse Response (IIR) estimation, comparison with Least-Squares



Estimates within two standard deviations



Magenta: noisy data; blue: noise-free data

Example 2: Direct data-driven control

$$
\begin{aligned}
\min_{\hat{\mathrm{u}},\hat{\mathrm{y}}} \quad & \sum_{k=0}^{L_s-1} \left(\left\| \hat{y}_k - y_k^{\mathrm{ref}} \right\|_Q^2 + \left\| \hat{u}_k - u_k^{\mathrm{ref}} \right\|_R^2 \right) \\
\text{s.t.} \quad & \left(\begin{bmatrix} \mathrm{u}_{\mathrm{ini}} \\ \hat{\mathrm{u}} \end{bmatrix}, \begin{bmatrix} \mathrm{y}_{\mathrm{ini}} \\ \hat{\mathrm{y}} \end{bmatrix} \right) \text{ is a trajectory,} \\
& \hat{\mathrm{u}} \in \mathcal{U}, \hat{\mathrm{y}} \in \mathcal{Y},
\end{aligned}
$$

$$
\hat{\mathrm{z}} = \begin{bmatrix} \mathrm{u}_{\mathrm{ini}} \\ \mathrm{u}^{\mathrm{ref}} \\ \mathrm{y}_{\mathrm{ini}} \\ \mathrm{y}^{\mathrm{ref}} \end{bmatrix}, \quad \mathrm{z} = \begin{bmatrix} 0 \\ \hat{\mathrm{u}} \\ 0 \\ \hat{\mathrm{y}} \end{bmatrix}, \quad
\begin{aligned}
\Sigma_{u_i} &= \sigma_{p,u}^2 I_m \otimes I_{L_s} \\
\Sigma_{u_s} &= R^{-1} \otimes I_{L_s} \\
\Sigma_{y_i} &= \sigma_{p,y}^2 I_p \otimes I_{L_s} \\
\Sigma_{y_s} &= Q^{-1} \otimes I_{L_s}
\end{aligned}
$$

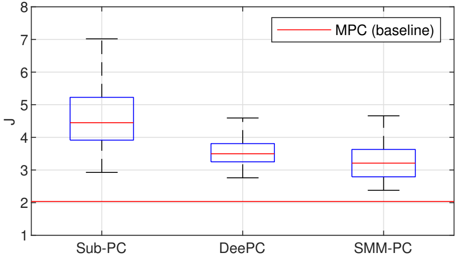
Example 2: Direct data-driven control

$$\min_{\hat{u},\hat{y}} \sum_{k=0}^{L_s-1} \left(\left\| \hat{y}_k - y_k^{\text{ref}} \right\|_Q^2 + \left\| \hat{u}_k - u_k^{\text{ref}} \right\|_R^2 \right)$$

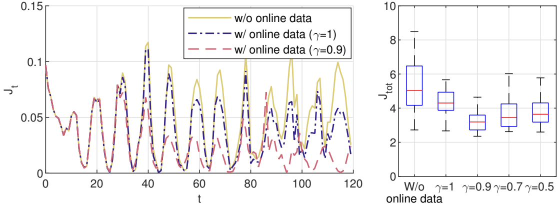
$$\text{s.t.} \quad \left(\begin{bmatrix} u_{\text{ini}} \\ \hat{u} \end{bmatrix}, \begin{bmatrix} y_{\text{ini}} \\ \hat{y} \end{bmatrix} \right) \text{ is a trajectory,}$$

$$\hat{u} \in \mathcal{U}, \hat{y} \in \mathcal{Y},$$

$$\hat{z} = \begin{bmatrix} u_{\text{ini}} \\ u^{\text{ref}} \\ y_{\text{ini}} \\ y^{\text{ref}} \end{bmatrix}, \quad z = \begin{bmatrix} 0 \\ \hat{u} \\ 0 \\ \hat{y} \end{bmatrix}, \quad \begin{aligned} \Sigma_{u_i} &= \sigma_{p,u}^2 I_m \otimes I_{L_s} \\ \Sigma_{u_s} &= R^{-1} \otimes I_{L_s} \\ \Sigma_{y_i} &= \sigma_{p,y}^2 I_p \otimes I_{L_s} \\ \Sigma_{y_s} &= Q^{-1} \otimes I_{L_s} \end{aligned}$$



CL cost comparison with *competitors*



Adaptation using online data

Confidence regions for data-driven prediction

What is the error associated with predicting y_s from noisy data?

Confidence regions for data-driven prediction

What is the error associated with predicting y_s from noisy data?

Prototype of stochastic data-driven predictors

Existing algorithms [a] can be expressed as

$$\mathcal{F}_Z(u_{ini}, u_s, y_{ini}) = Y_f \operatorname*{argmin}_{g(,\delta)} \|\delta\|_2^2 + \lambda \|g\|_2^2 \quad \text{s.t.} \quad \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g = \begin{bmatrix} u_{ini} \\ u_s \\ y_{ini} + \delta \end{bmatrix} \tag{PRED}$$

where λ is algorithm-dependent.

[a] Subspace predictor [Fiedler, 2021], SMM, Wasserstein distance minimization [Lian, 2021]

Goal: obtain **probabilistic** and **deterministic** confidence regions

SMM as data-driven predictor

- Output noise ($\mathrm{y_{ini}}$, Y_p, Y_f)
- Specific matrix structure: M length-L input-output trajectories

$$z_i^d = \mathrm{col}\left(u_{t_i}^d, \cdots, u_{t_i+L-1}^d, y_{t_i}^d, \cdots, y_{t_i+L-1}^d\right) \in \mathbb{R}^{L(n_u+n_y)}, \quad i = 0, \cdots, M-1$$

Matrix $Z = \begin{bmatrix} z_0^d & \cdots & z_{M-1}^d \end{bmatrix}$ is mosaic Hankel $(t_{i+1} = t_i + 1)$ or Page $(t_{i+1} = t_i + L)$

SMM as data-driven predictor

- Output noise (y_{ini}, Y_p, Y_f)
- Specific matrix structure: M length-L input-output trajectories

$$z_i^d = \text{col}\left(u_{t_i}^d, \cdots, u_{t_i+L-1}^d, y_{t_i}^d, \cdots, y_{t_i+L-1}^d\right) \in \mathbb{R}^{L(n_u+n_y)}, \quad i = 0, \cdots, M-1$$

Matrix $Z = \begin{bmatrix} z_0^d & \cdots & z_{M-1}^d \end{bmatrix}$ is mosaic Hankel $(t_{i+1} = t_i + 1)$ or Page $(t_{i+1} = t_i + L)$

The ML estimate of g solves

$$\boxed{\min_g \text{ logdet}\left(\begin{bmatrix} \Sigma_{Y_p}(g) & 0 \\ 0 & \Sigma_{Y_f}(g) \end{bmatrix} + \begin{bmatrix} \Sigma_{y_i} & 0 \\ 0 & 0 \end{bmatrix}\right) + \delta^\top \left(\begin{bmatrix} \Sigma_{Y_p}(g) & 0 \\ 0 & \Sigma_{Y_f}(g) \end{bmatrix} + \Sigma_{y_i}\right)^{-1} \underbrace{\delta}_{Y_p g - y_{ini}},}$$

where $\Sigma_{y_i} = \sigma_{p,y}^2 I_{n_y L_0}$, $\Sigma_{Y_p}(g) = \sigma_{p,y}^2 \|g\|_2^2 I_{n_y L_0}$, $\Sigma_{Y_f}(g) = \sigma_y^2 \|g\|_2^2 I_{n_y L_s}$.

SMM as data-driven predictor

- Output noise ($\mathrm{y_{ini}}$, Y_p, Y_f)
- Specific matrix structure: M length-L input-output trajectories

$$z_i^d = \mathrm{col}\left(u_{t_i}^d, \cdots, u_{t_i+L-1}^d, y_{t_i}^d, \cdots, y_{t_i+L-1}^d\right) \in \mathbb{R}^{L(n_u+n_y)}, \quad i = 0, \cdots, M-1$$

Matrix $Z = \begin{bmatrix} z_0^d & \cdots & z_{M-1}^d \end{bmatrix}$ is mosaic Hankel $(t_{i+1} = t_i + 1)$ or Page $(t_{i+1} = t_i + L)$

The ML estimate of g solves

$$\boxed{\min_g \ \mathrm{logdet}\left(\begin{bmatrix} \Sigma_{Y_p}(g) & 0 \\ 0 & \Sigma_{Y_f}(g) \end{bmatrix} + \begin{bmatrix} \Sigma_{y_i} & 0 \\ 0 & 0 \end{bmatrix}\right) + \delta^\top \left(\begin{bmatrix} \Sigma_{Y_p}(g) & 0 \\ 0 & \Sigma_{Y_f}(g) \end{bmatrix} + \Sigma_{y_i}\right)^{-1} \underbrace{\delta}_{Y_p g - y_{ini}},}$$

where $\Sigma_{y_i} = \sigma_{p,y}^2 I_{n_y L_0}$, $\Sigma_{Y_p}(g) = \sigma_{p,y}^2 \|g\|_2^2 I_{n_y L_0}$, $\Sigma_{Y_f}(g) = \sigma_y^2 \|g\|_2^2 I_{n_y L_s}$.

Quadratic approximation of the obj. about $g = \hat{g}$ gives (PRED) with $\lambda = n_y \left(L_0 \sigma_{p,y}^2 / \|\hat{g}\|_2^2 + L\sigma_y^2\right)$.

Confidence regions for data-driven predictors

Probabilistic and robust (in probability) bounds for PRED

Consider prediction $\mathrm{y} = \mathcal{F}_Z(\cdot)$ of true output y_0. Then

$$\mathrm{y} - \mathrm{y}_0 \mid g, \delta \sim \mathcal{N}\left(\Gamma\delta, \underbrace{\begin{bmatrix} -\Gamma & I_{n_y L'} \end{bmatrix} \begin{bmatrix} \Sigma_{Y_p}(g) & 0 \\ 0 & \Sigma_{Y_f}(g) \end{bmatrix} \begin{bmatrix} -\Gamma^\top \\ I_{n_y L'} \end{bmatrix} + \Gamma\, \Sigma_{y_i} \Gamma^\top}_{\Sigma}\right)$$

where $\Gamma = \operatorname{col}\left(CA^{L_0}, \cdots, CA^{L-1}\right) \operatorname{col}\left(C, \cdots, CA^{L_0-1}\right)^\dagger$.

Confidence regions for data-driven predictors

Probabilistic and robust (in probability) bounds for PRED

Consider prediction $\mathrm{y} = \mathcal{F}_Z(\cdot)$ of true output y_0. Then

$$\mathrm{y} - \mathrm{y}_0 \mid g, \delta \sim \mathcal{N}\left(\Gamma\delta, \underbrace{\begin{bmatrix} -\Gamma & I_{n_y L'} \end{bmatrix} \begin{bmatrix} \Sigma_{Y_p}(g) & 0 \\ 0 & \Sigma_{Y_f}(g) \end{bmatrix} \begin{bmatrix} -\Gamma^\top \\ I_{n_y L'} \end{bmatrix} + \Gamma\, \Sigma_{y_i} \Gamma^\top}_{\Sigma} \right)$$

where $\Gamma = \operatorname{col}\left(CA^{L_0}, \cdots, CA^{L-1}\right) \operatorname{col}\left(C, \cdots, CA^{L_0-1}\right)^\dagger$.

The true output y_0 belongs w.p. p to

$$\mathcal{Y} = \left\{ \tilde{y} \mid (\mathrm{y} - \tilde{y} - \Gamma\delta)^\mathsf{T} \Sigma^{-1} (\mathrm{y} - \tilde{y} - \Gamma\delta) \leq \mu_p \right\}, \quad F_{\chi^2(L')}(\mu_p) = p$$

$F_{\chi^2(d)}(\cdot)$ is the cumulative distribution function of the χ^2-distribution with d degrees of freedom.

Insights from the proof

- $$\mathrm{y} - \mathrm{y}_0 = \underbrace{E_f g}_{\text{noise in } Y_f} + \underbrace{y^-}_{\text{wrong initial condition}}$$

  $E_f = Y_f - Y_f^0; \quad y^-$ free response from I.C. $u_{\text{ini}}^- = 0,\ y_{\text{ini}}^- = Y_p^0 g - y_{\text{ini}}^0$

Insights from the proof

- $\mathrm{y} - \mathrm{y}_0 = \underbrace{E_f g}_{\text{noise in } Y_f} + \underbrace{y^-}_{\text{wrong initial condition}}$

  $E_f = Y_f - Y_f^0$; $\quad y^-$ free response from I.C. $u_{\text{ini}}^- = 0$, $y_{\text{ini}}^- = Y_p^0 g - y_{\text{ini}}^0$
- $\mathrm{y}_{\text{ini}}^- = \delta + (\mathrm{y}_{\text{ini}} - \mathrm{y}_{\text{ini}}^0) - E_p g$

$$\mathrm{y}^- = \begin{bmatrix} CA^{L_0} \\ \vdots \\ CA^{L-1} \end{bmatrix} x^-, \ \mathrm{y}_{\text{ini}}^- = \begin{bmatrix} C \\ \vdots \\ CA^{L_0-1} \end{bmatrix} x^- \Rightarrow \mathrm{y}^- = \Gamma\, \mathrm{y}_{\text{ini}}^-$$

Insights from the proof

- $\mathrm{y} - \mathrm{y}_0 = \underbrace{E_f g}_{\text{noise in } Y_f} + \underbrace{y^-}_{\text{wrong initial condition}}$

  $E_f = Y_f - Y_f^0$; $\quad y^-$ free response from I.C. $u_{\text{ini}}^- = 0$, $y_{\text{ini}}^- = Y_p^0 g - y_{\text{ini}}^0$
- $\mathrm{y}_{\text{ini}}^- = \delta + (\mathrm{y}_{\text{ini}} - \mathrm{y}_{\text{ini}}^0) - E_p g$

$$\mathrm{y}^- = \begin{bmatrix} CA^{L_0} \\ \vdots \\ CA^{L-1} \end{bmatrix} x^-, \ \mathrm{y}_{\text{ini}}^- = \begin{bmatrix} C \\ \vdots \\ CA^{L_0-1} \end{bmatrix} x^- \Rightarrow \mathrm{y}^- = \Gamma \, \mathrm{y}_{\text{ini}}^-$$

- Holds for general predictors

$$\mathcal{F}_Z(\cdot) = Y_f g, \quad \text{s.t.} \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g = \begin{bmatrix} \mathrm{u}_{\text{ini}} \\ \mathrm{u}_{\text{s}} \\ \mathrm{y}_{\text{ini}} + \delta \end{bmatrix}$$

Observations

✗ Confidence regions depend on **model's quantities** (range space of observability); no closed form solution in more general (noise and matrix structures) cases

Observations

✗ Confidence regions depend on **model's quantities** (range space of observability); no closed form solution in more general (noise and matrix structures) cases

✓ Quantification of the prediction error due to noisy information; tool for designing **new predictors**

Minimum mean-squared error predictor

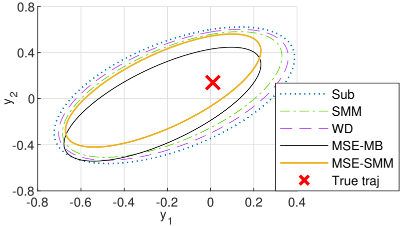
The minimum MSE predictor of the class (PRED) is

$$\mathcal{F}_Z(\cdot) = Y_f \operatorname*{argmin}_{g(,\delta)} \underbrace{\delta^\top \Gamma^\top \Gamma \delta + \mathrm{tr}(\Sigma)}_{\mathrm{MSE}(g,\delta)} = \operatorname*{argmin}_{g(,\delta)} \; \|\delta\|_Q^2 + \lambda_{\mathrm{MSE}} \|g\|_2^2$$

$$\text{s.t.} \quad \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g = \begin{bmatrix} \mathrm{u}_{\mathrm{ini}} \\ \mathrm{u}_{\mathrm{s}} \\ \mathrm{y}_{\mathrm{ini}} + \delta \end{bmatrix} \tag{MSE-PRED}$$

Numerical experiments

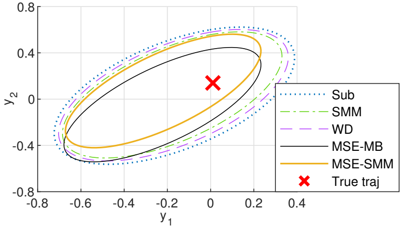
Two-output 4th order system. $L = 10$, $L_0 = 8$, $M = 80$, $\sigma^2 = 0.1$, $p = 0.90$, unit Gaussian input.



CR of different predictors (based on true Γ).

Numerical experiments

Two-output 4th order system. $L = 10$, $L_0 = 8$, $M = 80$, $\sigma^2 = 0.1$, $p = 0.90$, unit Gaussian input.



CR of different predictors (based on true Γ).

1000 randomly generated systems (states between 3 and 8). $L = 20$, $L_0 = 8$, $L' = 12$, $M = 320$, unit Gaussian input.

Comparison of the empirical MSE.

	$\sigma^2 = 0.1$	$\sigma^2 = 0.5$	$\sigma^2 = 1$
Sub	0.115	0.558	1.106
SMM	0.099	0.476	0.915
WD	0.113	0.548	1.091
MSE-SMM	0.096	0.460	0.897
MSE-MB	0.094	0.435	0.833

Data informativity in direct data-driven simulation

How to best choose u_{d}?

- The *Fundamental Lemma* is effectively an **experiment design** result
- Goal: go beyond PRBS (Pseudo Random Binary Signals) and randomly selected input..
- **Data informativity** for the *noisy* given-input simulation problem
- Informativity encompasses both the selection of the input and the matrix structure (Hankel/Page)

Bayesian design of experiment (DOE)

Interpret the output trajectory y_s as a random variable.

$$y_s^1 \sim \mathcal{N}(0, \Sigma_K) \quad \rightarrow \quad \textit{Prior distribution}$$

$$y_s^2 \sim \mathcal{N}(Y_f g^*, \quad \underbrace{\Sigma_{Y_f}(g^*)}_{\text{sub-matrix of } \Sigma_z(g^*)} \quad) \quad \rightarrow \quad \textit{Data-based estimate (via SMM)}$$

Bayesian design of experiment (DOE)

Interpret the output trajectory y_s as a random variable.

$$y_\mathrm{s}^1 \sim \mathcal{N}(0, \Sigma_\mathrm{K}) \quad \rightarrow \quad \textit{Prior distribution}$$

$$y_\mathrm{s}^2 \sim \mathcal{N}(Y_f g^*, \underbrace{\Sigma_{Y_f}(g^*)}_{\text{sub-matrix of } \Sigma_z(g^*)}) \quad \rightarrow \quad \textit{Data-based estimate (via SMM)}$$

Combine prior and data-based estimates to provide the posterior

$$(y_\mathrm{s}|(u_\mathrm{d}, y_\mathrm{d})) \sim \mathcal{N}(K(Y_f g^*), \Sigma_\mathrm{post}) \quad \rightarrow \quad \textit{Posterior of } y_\mathrm{s}$$

where $K = \Sigma_\mathrm{K}(\Sigma_\mathrm{K} + \Sigma_{Y_f})^{-1}$ is the Kalman gain and $\Sigma_\mathrm{post} = \Sigma_\mathrm{K} - \Sigma_\mathrm{K}(\Sigma_\mathrm{K} + \Sigma_{Y_f})^{-1}\Sigma_\mathrm{K}$ is the posterior covariance.

The experiment is **meaningful** when the *difference* between prior and posterior distributions is **maximized**.

Mutual information and information criterion

Mutual information [Cover, 1991] of two multivariate random variables x and y

$$I(x;y) = H(x) + H(y) - H(x,y) = H(x) - H(x|y)$$

$$\Rightarrow \; I(\mathrm{y}_\mathrm{s};(u_\mathrm{d}, y_\mathrm{d})) = \mathbb{E}\left[\underbrace{D\left(\mathrm{y}_\mathrm{s}|(u_\mathrm{d}, y_\mathrm{d})||p(\mathrm{y}_\mathrm{s})\right)}_{\text{KL divergence between prior and posterior trajectory}}\right]$$

Mutual information and information criterion

Mutual information [Cover, 1991] of two multivariate random variables x and y

$$I(x;y) = H(x) + H(y) - H(x,y) = H(x) - H(x|y)$$

$$\Rightarrow \quad I(\mathrm{y_s};(u_\mathrm{d},y_\mathrm{d})) = \mathbb{E}\left[\underbrace{D\left(\mathrm{y_s}|(u_\mathrm{d},y_\mathrm{d})||p(\mathrm{y_s})\right)}_{\text{KL divergence between prior and posterior trajectory}}\right]$$

Information criterion for SMM

If $\Sigma_{Y_f}(g) = \sigma_y^2 \|g\|_2^2 \, I_{n_y L_s}$, then there exists a function f such that

$$I(y_\mathrm{s};(u_\mathrm{d},y_\mathrm{d})) = f(\|g\|_2^2, \Sigma_\mathrm{K})$$

where $f(\|g\|_2^2, \cdot)$ is monotonically decreasing.

Bi-level input design problem

The input data trajectory can be designed to maximize the information:

$$u_{\mathrm{d}\ [0,N-1]} \rightarrow U, Y \rightarrow g^* \rightarrow I$$

Bi-level input design problem

The input data trajectory can be designed to maximize the information:

$$u_{\mathrm{d}\,[0,N-1]} \to U, Y \to g^* \to I$$

$$\min_{g,u_\mathrm{d}} \|g\|_2^2 \quad \to \quad \textit{information criterion}$$

$$\text{s.t.} \quad u_\mathrm{d} \in \mathcal{U} \quad \to \quad \textit{input constraints}$$

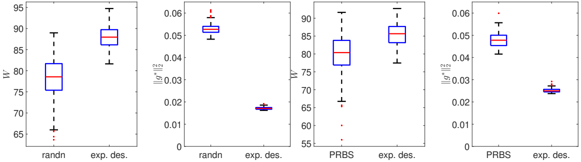
$$g \in \arg\min_{g \in \mathcal{G}, u_\mathrm{d}} \; \mathrm{logdet}(\Sigma_\mathrm{y}(g)) + \begin{bmatrix} Y_\mathrm{p} g - y_\mathrm{ini} \\ 0 \end{bmatrix}^\top \Sigma_\mathrm{y}^{-1}(g) \begin{bmatrix} Y_\mathrm{p} g - y_\mathrm{ini} \\ 0 \end{bmatrix} \to \quad \textit{ML estimator}$$

Two **challenges** associated with the bi-level program:

- Inner-problem depends on Y_p (implicitly function of u_d)
- Dealing with the bi-level program

Truncated IIR estimation with Hankel matrices

Fit metric: $W = 100\left(1 - \left[\frac{\sum_{i=0}^{L_s-1}(y_{s,i} - \hat{y}_{s,i})^2}{\sum_{i=0}^{L_s-1}(y_{s,i} - \bar{y}_s)^2)^2}\right]^{1/2}\right)$

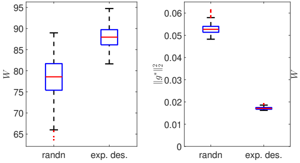


Energy constraint. Magnitude constraint.

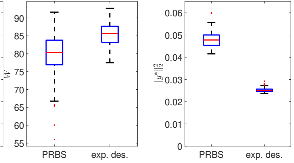
- Optimally designed input perform better than standard PE inputs
- Informativity key to achieve the improvement

Truncated IIR estimation with Hankel matrices

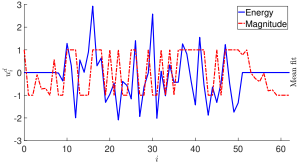
Fit metric: $W = 100\left(1 - \left[\frac{\sum_{i=0}^{L_s-1}(y_{s,i} - \hat{y}_{s,i})^2}{\sum_{i=0}^{L_s-1}(y_{s,i} - \bar{y}_s)^2)^2}\right]^{1/2}\right)$



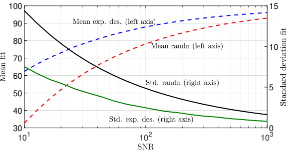
Energy constraint.



Magnitude constraint.



Optimal inputs.



Fit as a function of SNR.

- Optimally designed input perform better than standard PE inputs
- Informativity key to achieve the improvement
- Optimal input do not have “intuitive” shape
- Impact of DOE is a function of the SNR

Take-away points

- Statistical framework for the direct prediction and control problems
- First prediction error bounds and data informativity studies in this implicit non-parametric setting
- Towards a principled way to reason about robustness in data-driven problems

References and open-access repositories (with citable DOI) on the ETH Research Collection portal https://www.research-collection.ethz.ch.

* Yin, AI, Smith "Maximum Likelihood Estimation in Data-Driven Modeling and Control", IEEE TAC (Early Access) 2021
  code repository: https://doi.org/10.3929/ethz-b-000525037
* Yin, AI, Smith "Maximum Likelihood Signal Matrix Model for Data-Driven Predictive Control", Learning for Dynamic and Control (L4DC) 2021
  code repository: https://doi.org/10.3929/ethz-b-0005140019
* Yin, AI, Smith "Data-driven prediction with stochastic data: confidence regions and minimum mean-squared error estimates", European Control Conference (ECC) 2022; code repository: available soon!
* AI, Yin, Smith "Experiment design for impulse response identification with signal matrix models", IFAC Symposium on System Identification (SYSID) 2021
  code repository: https://doi.org/10.3929/ethz-b-000511316
* AI, Yin, Smith "Design of input for data-driven simulation with Hankel and Page matrices", IEEE Conference on Decision and Control (CDC) 2021
  code repository: https://doi.org/10.3929/ethz-b-000514080

Outlook

- Stochastic MPC formulation based on prediction error-bounds
- Informative input for control (dual control in direct problems)**
- Adaptive signal matrices: when and why?

** AI, Smith "A multiobjective LQR synthesis approach to dual control for uncertain plants", IEEE Control Systems Letters, 2020
code repository: https://doi.org/10.3929/ethz-b-000472064

** Parsi, AI, Smith "Active exploration in adaptive model predictive control", IEEE Conference on Decision and Control (CDC) 2020
code repository: https://doi.org/10.5905/ethz-1007-430

** Parsi, AI, Smith "An explicit dual control approach for constrained reference tracking of uncertain linear systems", IEEE TAC (accepted), 2022
code repository: https://doi.org/10.3929/ethz-b-000518264

References

- A note on persistency of excitation
  Willems, Rapisarda, Markovsky, de Moor. *Systems & Control Letters* (2005)
- From time series to linear system–Part I, Part II, Part III
  Willems. *Automatica* (1987)
- Data-driven simulation and control
  Markovsky, Rapisarda. *International Journal of Control* (2008)
- Adaptive robust data driven building control via bi-level reformulation: an experimental result
  Lian, Shi, Koch, Jones. *Arxiv* (2021)
- On the relationship between data-enabled predictive control and subspace predictive control
  Fiedler, Lucia. *European Control Conference* (2021)
- Elements of Information Theory
  Cover, Thomas. *Wiley-Interscience* (1991)

INSTITUT FÜR AUTOMATIK

Andrea Iannelli
iannelli@control.ee.ethz.ch
http://people.ee.ethz.ch/~andreaia

Thank you for your attention!