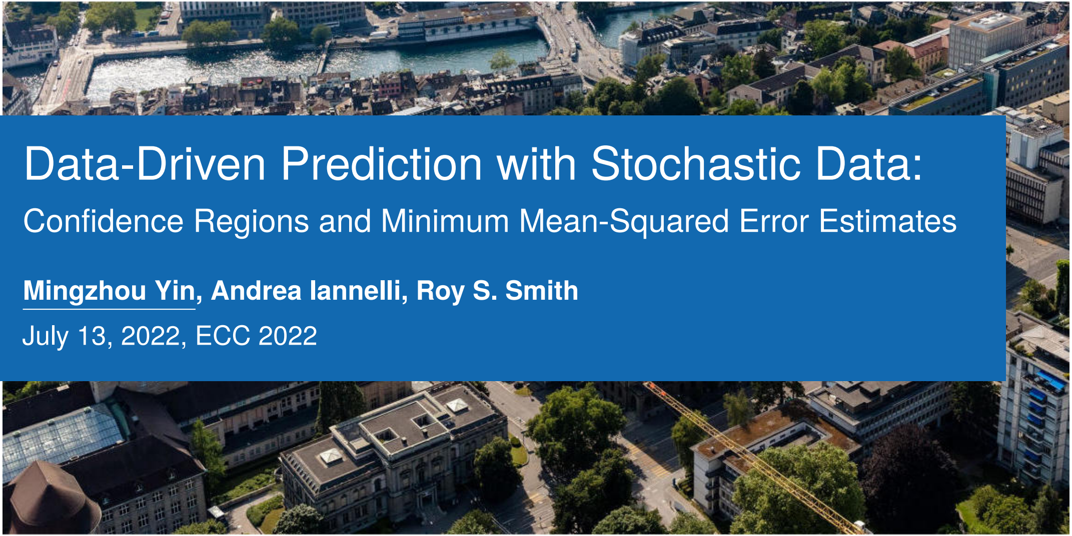
ETH zürich

Automatic Control Laboratory



Data-Driven Prediction with Stochastic Data:

Confidence Regions and Minimum Mean-Squared Error Estimates

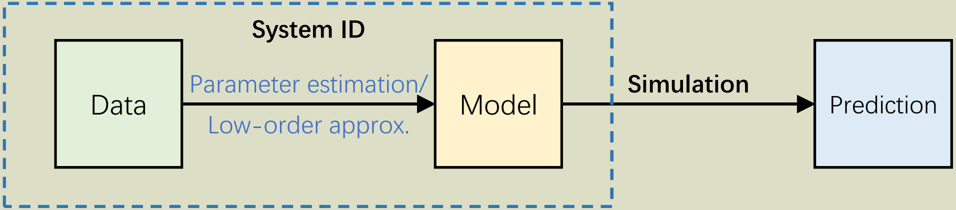
Mingzhou Yin, Andrea Iannelli, Roy S. Smith

July 13, 2022, ECC 2022

Towards Data-Driven Trajectory Prediction

- Fundamental in simulation, analysis, and **predictive control**
- Conventional paradigm relies on **models**
- ... to parameterize system knowledge compactly

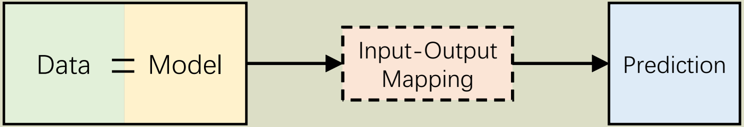
Paradigm of system identification



Towards Data-Driven Trajectory Prediction

- For complex systems, however: modeling is hard, but data is big
- **Willems' fundamental lemma**: a big data matrix can characterize all trajectories of a given horizon for linear systems

Paradigm of data-driven prediction



- ... in the noise-free case
- Only approximate solution for stochastic data $\rightarrow$ a **stochastic** predictor needed

Problem Statement

- Discrete-time LTI system with output noise
- Input-output trajectory data collected in a **signal matrix**

$$Z = \begin{bmatrix} z_0^d & \cdots & z_{M-1}^d \end{bmatrix}, \quad z_i^d = \text{col}\left(u_{t_i}^d, \cdots, u_{t_i+L-1}^d, y_{t_i}^d, \cdots, y_{t_i+L-1}^d\right)$$

- ... either Hankel ($t_{i+1} = t_i + 1$), Page ($t_{i+1} = t_i + L$), or multiple experiments
- **Objective:** Find the input-output mapping using only Z

$$\underbrace{\begin{bmatrix} y_0 \\ \vdots \\ y_{L'-1} \end{bmatrix}}_{\mathbf{y}} = \mathcal{F}_Z\left(\underbrace{\begin{bmatrix} u_0 \\ \vdots \\ u_{L'-1} \end{bmatrix}}_{\mathbf{u}}, \underbrace{\begin{bmatrix} u_{-L_0} \\ \vdots \\ u_{-1} \end{bmatrix}}_{\mathbf{u}_{\text{ini}}}, \underbrace{\begin{bmatrix} y_{-L_0} \\ \vdots \\ y_{-1} \end{bmatrix}}_{\mathbf{y}_{\text{ini}}}\right), \quad L = L_0 + L'$$

A Battle Against Noise

- Define partition $Z = \text{col}\,(U_p, U_f, Y_p, Y_f)$
- Noise-free solution:

$$\mathbf{y} = Y_f g, \quad g : \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g = \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{u} \\ \mathbf{y}_{\text{ini}} \end{bmatrix}$$

- ... not well-defined for noisy case

- Output noise case: need to fix a unique g
- Difficulty: stochasticity in $\mathbf{y}_{\text{ini}}$, Y_p, Y_f

The Prototype Solution

$$\mathcal{F}_Z(\cdot) = Y_f \operatorname*{argmin}_g \quad \|\delta\|_2^2 + \lambda \|g\|_2^2$$

$$\text{s.t.} \quad \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g = \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{u} \\ \mathbf{y}_{\text{ini}} + \delta \end{bmatrix}$$

- $\|\delta\|_2^2$: deviation from past output measurements
- $\|g\|_2^2$: uncertainty of prediction
- Different λ choices proposed:
  - Prediction error method / subspace predictor (*Sub*): $\lambda \to 0$
  - Maximum likelihood signal matrix model (*SMM*): $\lambda = n_y \left(L'\sigma^2 / \|g_{\text{pinv}}\|_2^2 + L\sigma^2 \right)$
  - Wasserstein distance minimization (*WD*): $\lambda = n_y L_0 \sigma^2$

Quantify Prediction Error

- Essential in predictive control with **robustness** and **safety** constraints
- ... but only loose bounds for bounded noise available

- **This work:** confidence region for general data-driven predictors
- Noise assumption: zero-mean Gaussian noise

$$\mathbf{y}_{\text{ini}} \sim \mathcal{N}\left(\mathbf{y}^0_{\text{ini}}, \Sigma_{\text{yini}}\right), \quad \text{vec}\left(\begin{bmatrix} Y_p \\ Y_f \end{bmatrix}\right) \sim \mathcal{N}\left(\text{vec}\left(\begin{bmatrix} Y_p^0 \\ Y_f^0 \end{bmatrix}\right), \Sigma_Y\right)$$

$$\begin{bmatrix} Y_p \\ Y_f \end{bmatrix} g \Big| g \sim \mathcal{N}\left(\begin{bmatrix} Y_p^0 \\ Y_f^0 \end{bmatrix} g, \underbrace{\begin{bmatrix} \Sigma_p & \Sigma_{pf} \\ \Sigma_{pf}^\mathsf{T} & \Sigma_f \end{bmatrix}}_{\Sigma_g}\right), \; \Sigma_g = \left(g^\mathsf{T} \otimes \mathbb{I}_{n_y L}\right) \Sigma_Y \left(g \otimes \mathbb{I}_{n_y L}\right)$$

Main Result

Theorem: Confidence region of stochastic data-driven predictors

The true output $\mathbf{y}_0$ is in the following ellipsoidal set w.p. p:

$$\mathcal{Y} = \left\{ \tilde{y} \mid (\mathbf{y} - \tilde{y} - \Gamma\delta)^{\mathsf{T}} \Sigma^{-1} (\mathbf{y} - \tilde{y} - \Gamma\delta) \leq \mu_p \right\}$$

where

$$\Sigma = \begin{bmatrix} -\Gamma & \mathbb{I}_{n_y L'} \end{bmatrix} \Sigma_g \begin{bmatrix} -\Gamma^{\mathsf{T}} \\ \mathbb{I}_{n_y L'} \end{bmatrix} + \Gamma \, \Sigma_{\mathsf{yini}} \Gamma^{\mathsf{T}}$$

Γ : autonomous transformation matrix from $\mathbf{y}_{\mathsf{ini}}$ to $\mathbf{y}$

$$\mu_p : F_{\chi^2(L')}(\mu_p) = p$$

Proof Sketch

- Two sources of error: $\mathbf{y} - \mathbf{y}_0 = \Gamma\left(\delta + \epsilon_{\text{ini}} - E_p g\right) + E_f g$.
- $\Gamma\left(\delta + \epsilon_{\text{ini}} - E_p g\right)$: error due to output initial condition mismatch

$$\left(Y_p^0 g - \mathbf{y}_{\text{ini}}^0\right) = \delta + \epsilon_{\text{ini}} - E_p g$$

- $E_f g$: error due to noise in Y_f

$$\Sigma = \underbrace{\begin{bmatrix} -\Gamma & \mathbb{I}_{n_y L'} \end{bmatrix} \Sigma_g \begin{bmatrix} -\Gamma^\mathsf{T} \\ \mathbb{I}_{n_y L'} \end{bmatrix}}_{\text{from } (-\Gamma E_p g + E_f g) \text{ term}} + \underbrace{\Gamma\, \Sigma_{\text{yini}} \Gamma^\mathsf{T}}_{\text{from } \Gamma \epsilon_{\text{ini}} \text{ term}}, \quad \Gamma\delta \text{ leads to the bias}$$

On Estimating Γ

- The confidence region depends on system parameter Γ
- ... but can be estimated by a data-driven approach

- Linear map $\Gamma \mathbb{d} = \mathcal{F}_Z(\mathbf{u} = \mathbf{0}, \mathbf{u}_{\text{ini}} = \mathbf{0}, \cdot)$
- Using the same data-driven predictor (and assume certainty equivalence)

$$\hat{\Gamma}_Z = Y_f \left(F^{-1} - F^{-1} U^\mathsf{T} (U F^{-1} U^\mathsf{T})^{-1} U F^{-1} \right) Y_p^\mathsf{T}, \quad F = \lambda \mathbb{I}_M + Y_p^\mathsf{T} Y_p$$

Beyond Confidence Region

- Mean-squared error can also be computed

$$\text{MSE}(g, \delta) = \delta^\mathsf{T} \Gamma^\mathsf{T} \Gamma \delta + \text{tr}\left(\begin{bmatrix} -\Gamma & \mathbb{I}_{n_y L'} \end{bmatrix} \Sigma_g \begin{bmatrix} -\Gamma^\mathsf{T} \\ \mathbb{I}_{n_y L'} \end{bmatrix} \right)$$

- Minimum MSE predictor

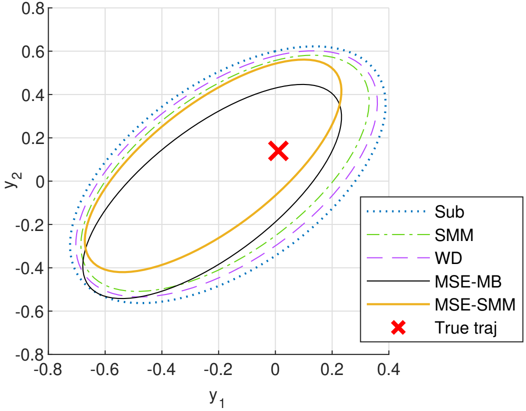
$$\mathcal{F}_Z(\cdot) = Y_f \operatorname*{argmin}_g \quad \text{MSE}(g, \delta)$$
$$\text{s.t.} \quad \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g = \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{u} \\ \mathbf{y}_{\text{ini}} + \delta \end{bmatrix}$$

Implications:

- Characterize the optimal data-driven predictor in terms of MSE
- Propose a new data-driven predictor by replacing Γ with $\hat{\Gamma}_Z$

A Toy Example

- Estimate the first two points in step response ($L' = 2$) of a fourth order system
- $L = 10$, $M = 80$, $\sigma^2 = 0.1$, $p = 0.9$
- Existing predictors perform similarly, *SMM* slightly better
- Min-MSE predictor with $\hat{\Gamma}_Z$ is close to the best possible one



Monte Carlo Campaign

- 1000 simulations with random systems of order 3 to 8
- $L = 20$, $L' = 12$, $M = 320$, random $\mathbf{u}, \mathbf{u}_{\text{ini}}, \mathbf{y}_{\text{ini}}$

Table: Empirical confidence levels

$p = 0.99$	CR-MB	CR-SMM
Sub	99.3%	99.8%
SMM	99.2%	99.7%
MSE-SMM	99.0%	99.2%

Table: 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-MB	0.094	0.435	0.833
MSE-SMM	0.096	0.460	0.897

A stochastic description of data-driven predictors

- Unified framework to analyze prediction errors for various data-driven predictors
- New data-driven predictor proposed based on the theoretically optimal predictor

AUTOMATIC CONTROL LABORATORY

Mingzhou Yin
myin@ethz.ch