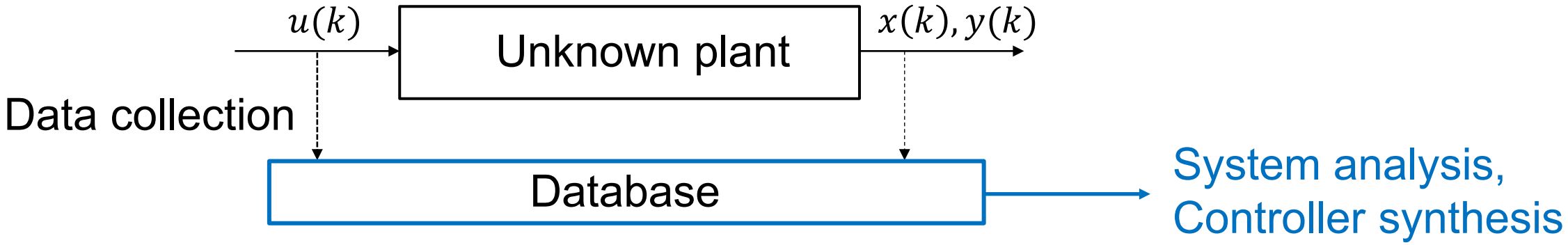
Data-Driven Control under Imperfect and Unreliable Data

Hideaki Ishii (Univ. of Tokyo)

Joint work with Iori Takaki and Ahmet Cetinkaya

Workshop on Trustworthy Data-Driven Control:
From Finite Samples to Robust Guarantees
IFAC WC, Busan, August 23rd, 2026

Background: Data-Driven Control

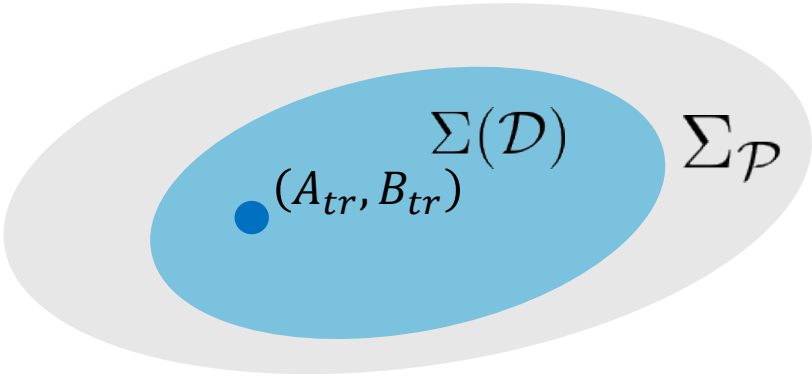


- Directly leverages measured data for analysis and controller synthesis.
  - No explicit model identification
- Our focus: Networked control under limited information / vulnerability
  - Quantized control
  - Cybersecurity

C. De Persis and P. Tesi (2020), *IEEE Trans. Autom. Control.*
J. Eising and J. Cortés (2022), *Proc. Eur. Control Conf.*
J. Jiao, H. J. van Waarde, H. L. Trentelman, M. K. Camlibel, and S. Hirche (2021), *Proc. IEEE Conf. Decis. Control.*

Data Informativity Framework

- $\mathcal{D}$ is informative for property $\mathcal{P}$:

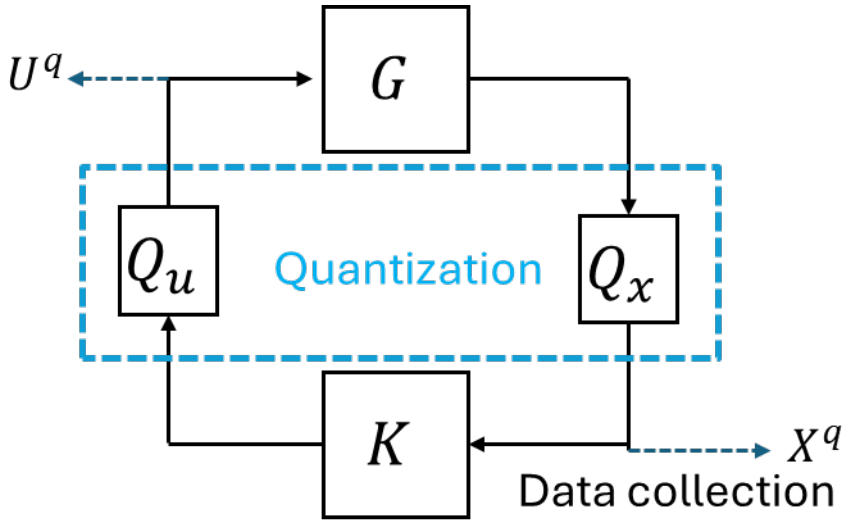


- Expressed via rank/LMI-type conditions on data matrices
- Under random noise, such conditions are hard to break.
- Informativity-based certification appears robust.

  — **What happens when data contain uncertainties or are even maliciously manipulated ?**

H. J. van Waarde, J. Eising, H. L. Trentelman, M. K. Camlibel (2020), *IEEE Trans. Autom. Control.*

1. Data-Driven Quantized Control



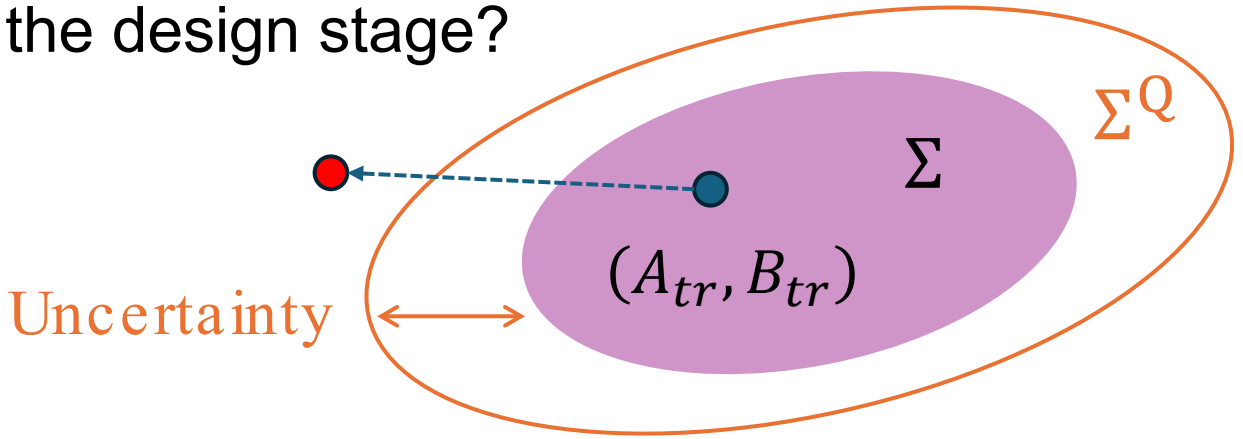
- Conventional data-driven control:

  Based on analog data

- Communication over digital channels:

  Quantized data

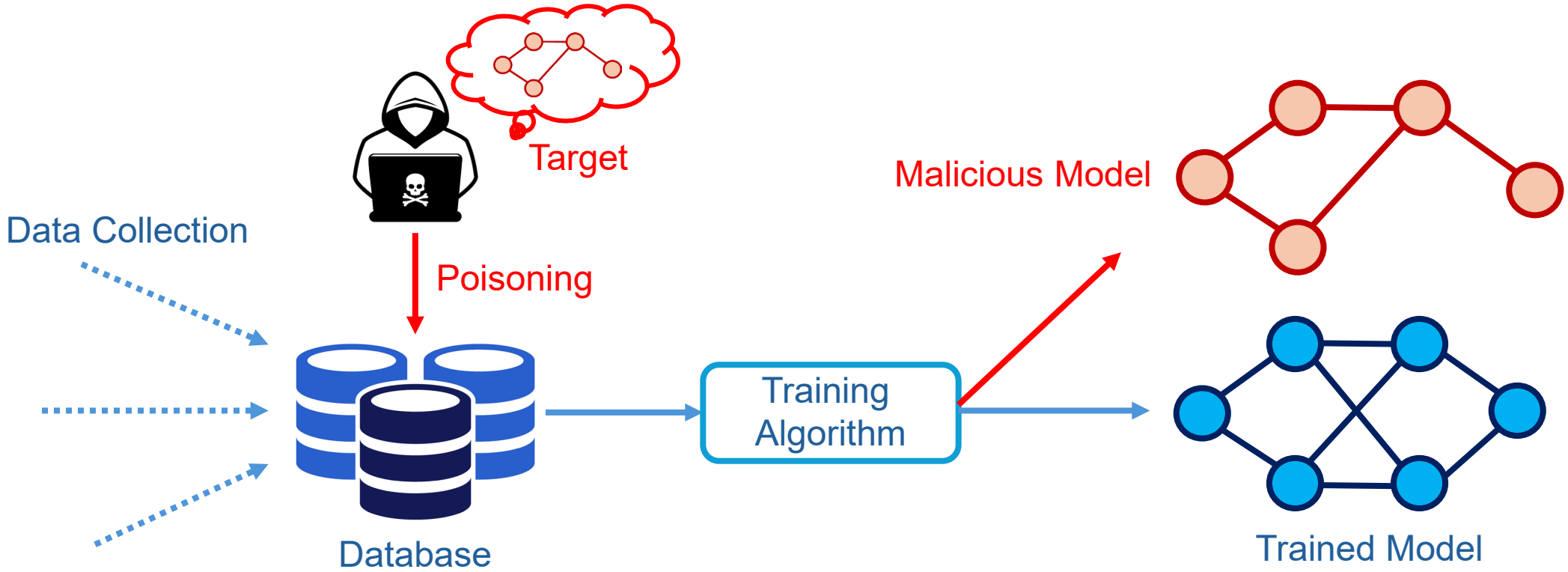
- System explained by the data becomes uncertain.
- If too coarse, the true system may be even indescribable.
- How coarse can quantization be at the design stage?



I. Takaki, A. Cetinkaya, and H. Ishii (2025), *Proc. IFAC NecSys.*

2. Data Poisoning Attacks on Informativity for Observability

- Malicious data are injected into the training dataset.



- Even small perturbations in the data can be enough to modify outcomes.

J. Fan, Q. Yan, M. Li, G. Qu, and Y. Xiao (2022), *IEEE Conf. Data Science in Cyberspace (DSC).*

Data Poisoning Attacks on Data-Driven Control

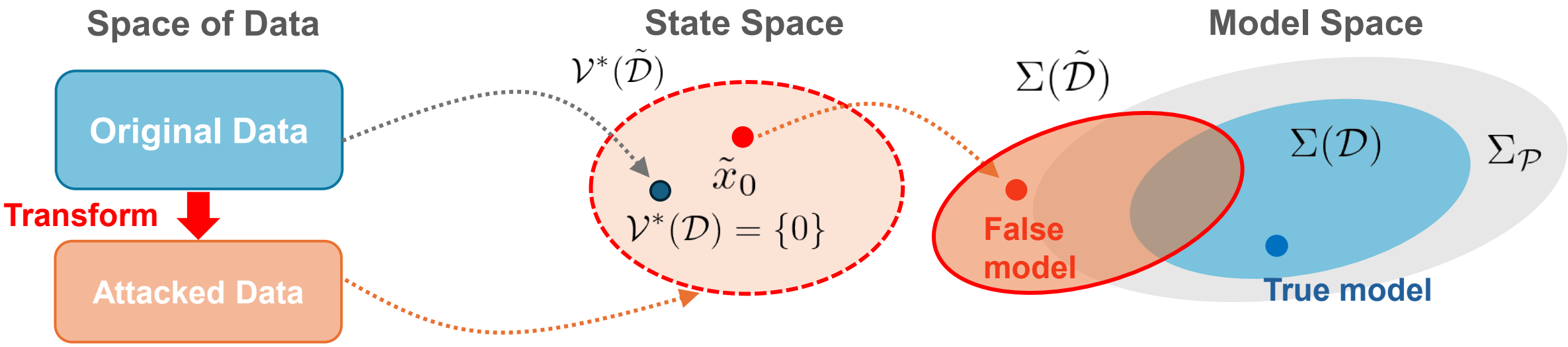
- In CPSs, adversaries can tamper with trajectory data of dynamical systems.

- Related works:

	Target	Attack Method	
Russo (2023)	LS estimate	Offline poisoning with stealth constraints	Small manipulations can alter estimates while remaining hard to detect statistically.
Sasahara (2026)	Data-driven LQR synthesis	Gradient-based adversarial perturbation	Small perturbations can destabilize synthesized controllers.
This work	Informativity for strong observability	Invertible block data transformations	Observability certificate can fail even when data rank is preserved.

A. Russo (2023), *Proc. Learning for Dynamics and Control Conf.*
H. Sasahara (2026), *Automatica.*

Poisoning Attacks Through Three Spaces



- **Focus of this work:** Malicious data transformations against informativity for strong observability.
- Invertible maps on time-series data induce transformations of the model set in a harmful manner.

  → To preserve richness in data (rank, PE)

Y. Tanaka and O. Kaneko (2024), *Kybernetika.*

System Model

- Consider the linear time-invariant system:

$$x_{k+1} = A_{\text{true}} x_k + B u_k + E w_k$$
$$y_k = C x_k + D u_k + F w_k$$

- $x_k \in \mathbb{R}^n$: state, $u_k \in \mathbb{R}^m$: input, $y_k \in \mathbb{R}^p$: output, $w_k \in \mathbb{R}^l$: noise

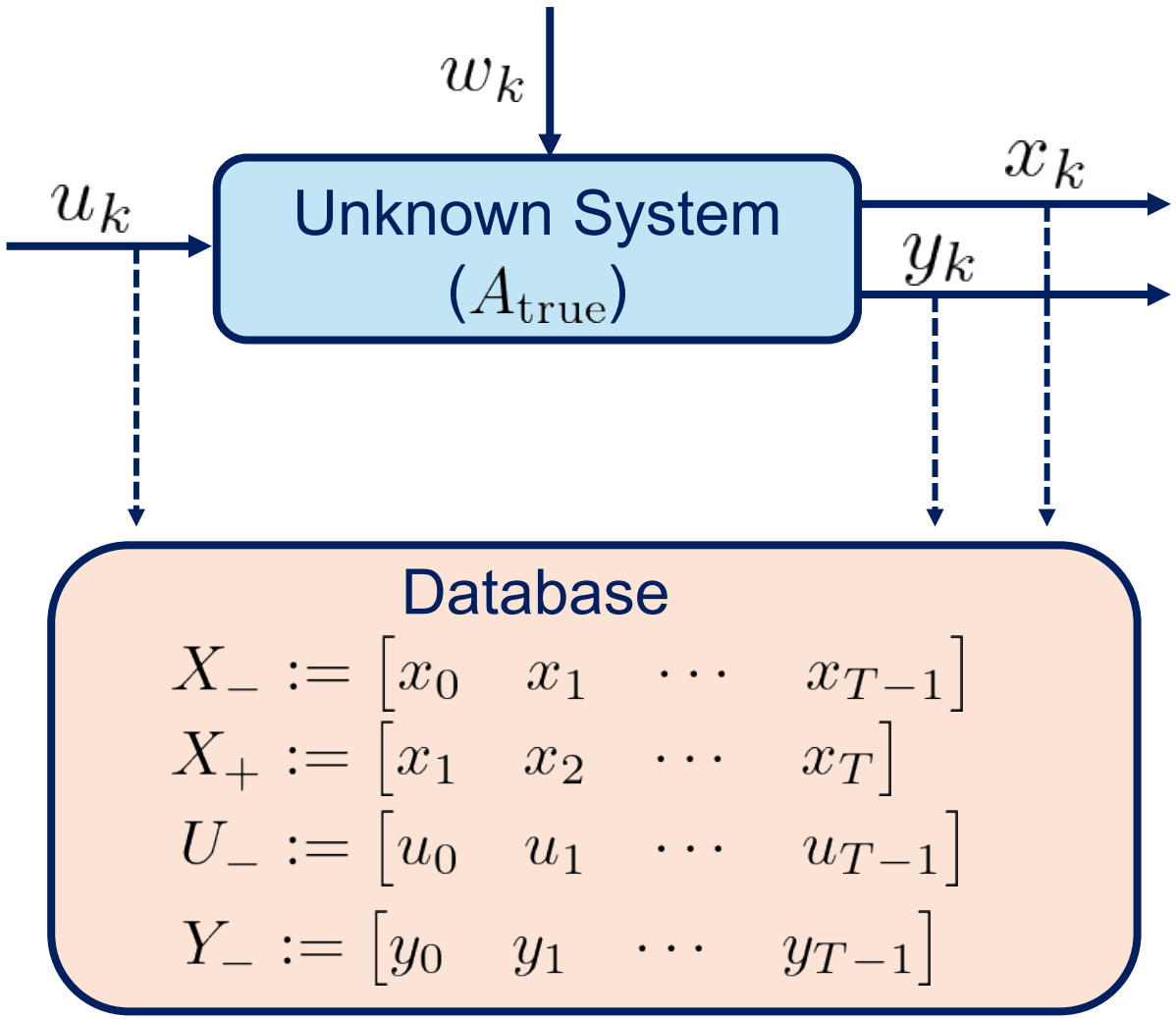
- A_{true} is unknown; B, C, D, E, and F are known.

Designer's goal:

Analyze system properties directly from measured data without identifying A_{true}.

Data Collection

- T: Data collection period
- $W_- := \begin{bmatrix} w_0 & w_1 & \cdots & w_{T-1} \end{bmatrix}$:
  Unknown noise sequence
- $\mathcal{D} := \begin{bmatrix} X_-^\mathsf{T} & X_+^\mathsf{T} & U_-^\mathsf{T} & Y_-^\mathsf{T} \end{bmatrix}^\mathsf{T} \in \mathbb{D}$:
  Stacked data



Noise Cancellation and the Affine Model Set

Assumption 1:

$$\exists M, N \text{ such that } \ker \begin{bmatrix} M & N \end{bmatrix} = \operatorname{im} \begin{bmatrix} E \\ F \end{bmatrix}$$

- Define matrices:

$$Q := M \qquad R(\mathcal{D}) := \begin{bmatrix} M & N \end{bmatrix} \begin{bmatrix} X_+ - BU_- \\ Y_- - CX_- - DU_- \end{bmatrix}$$

Affine model set:

$$\Sigma(\mathcal{D}) = \{A \in \mathbb{R}^{n \times n} |\ R(\mathcal{D}) = Q A X_-\}$$

- $\Sigma(\mathcal{D})$ contains all models consistent with data $\mathcal{D}$ after noise cancellation.

- Properties of all $A \in \Sigma(\mathcal{D})$ can be analyzed via data informativity.

Strong Observability

Definition 1

The system is called strongly observable if for each $x_0 \in \mathbb{R}^n$ and input sequence u, $y(k, x_0, u) = 0$ for all $k \in \mathbb{Z}_+$ implies $x_0 = 0$.

- Without input terms, this coincides with the usual observability.
- If the output remains zero under an input, the initial state must be zero.
- A key property for observer design and state estimation

Weakly Unobservable Subspace

Definition 2:

The weakly unobservable subspace $\mathcal{V}^*(A, B, C, D) \subseteq \mathbb{R}^n$ is defined as

$$\mathcal{V}^*(A, B, C, D) := \{x_0 \in \mathbb{R}^n | \ \exists u \text{ s.t. } y(k, x_0, u) = 0, \ k \in \mathbb{Z}_+\}.$$

- Strongly observable $\Leftrightarrow \mathcal{V}^*(A, B, C, D) = \{0\}$
- $\mathcal{V}^*$ is the largest output-nulling controlled invariant subspace:

$$\begin{bmatrix} A \\ C \end{bmatrix} \mathcal{V} \subseteq \mathcal{V} \times \{0\} + \text{im} \begin{bmatrix} B \\ D \end{bmatrix}$$

- This geometric characterization is fundamental for our attack design.

H. L. Trentelman, A. A. Stoorvogel, and M. Hautus (2001), *Springer.*

Coefficient Space Perspective

$v \in \mathbb{R}^T \quad \longrightarrow \quad x = X_- v$ $\qquad v$: Coefficient vector for combining the collected data

Data-driven geometric condition:

$$\begin{bmatrix} R(\mathcal{D}) \\ CX_- \end{bmatrix} \mathcal{J}(\mathcal{D}) \subseteq QX_- \mathcal{J}(\mathcal{D}) \times \{0\} + \operatorname{im} \begin{bmatrix} QB \\ D \end{bmatrix}$$

$$Q := M$$

$$R(\mathcal{D}) := \begin{bmatrix} M & N \end{bmatrix} \cdot \begin{bmatrix} X_+ - BU_- \\ Y_- - CX_- - DU_- \end{bmatrix}$$

Definition 3:

(i) $\mathcal{J}^*(\mathcal{D}) \subseteq \mathbb{R}^T$ — the maximum weakly unobservable coefficient space (= the largest $\mathcal{J}(\mathcal{D})$ satisfying the inclusion above)

(ii) $\mathcal{V}^*(\mathcal{D}) := X_- \mathcal{J}^*(\mathcal{D})$ — the data-driven weakly unobservable subspace

Behavioral interpretation: $\mathcal{J}(\mathcal{D})$ represents coefficients that generate unobservable trajectories from the collected data.

J. Eising and H. L. Trentelman (2021), *Syst. Control Lett.*

Informativity for Strong Observability

Definition 5:

The data $\mathcal{D}$ are informative for strong observability if $\Sigma(\mathcal{D}) \subseteq \Sigma_{\mathrm{SO}}$ (i.e., every model in the model set is strongly observable).

$$\Sigma_{\mathrm{SO}} := \{A \in \mathbb{R}^{n\times n} \mid (A, B, C, D) \text{ is strongly observable}\}$$

Lemma 1 [Eising and Trentelman (2021)]:

$\mathcal{D}$ are informative for strong observability if and only if:

(i) $C^{-1}\mathrm{im}\, D \subseteq \mathrm{im}\, P(\mathcal{D})$ (ii) $\mathcal{J}^*(\mathcal{D}) \subseteq \ker X_-$

- (i): Some output trajectories can be canceled through the input.
- (ii): Every coefficient in $\mathcal{J}^*(\mathcal{D})$ maps to zero state.

Our attack approach is to violate this condition.

J. Eising and H. L. Trentelman (2021), *Syst. Control Lett.*

Attack Model: Linear Transformation

- The attacker applies a block-diagonal linear transformation:

$$\Phi_a := \text{blk-diag}(\Phi_-^X, \Phi_+^X, \Phi_-^U, \Phi_-^Y)$$

- Attacked data: $\tilde{\mathcal{D}} := \Phi_a \mathcal{D}$

- Resulting in:

$$\tilde{X}_- := \Phi_-^X X_-, \ \tilde{X}_+ := \Phi_+^X X_+, \ \tilde{U}_- := \Phi_-^U U_-, \ \tilde{Y}_- := \Phi_-^Y Y_-$$

Key constraints:

- Φ_a must be nonsingular — rank-reducing attacks are easy to detect.
- Preserves data richness (rank, persistency of excitation)

Attack Goal: Destroying Informativity

- (By Lemma 1) $\tilde{\mathcal{D}}$ is NOT informative for strong observability if

$$\exists v \in \mathcal{J}^*(\tilde{\mathcal{D}}),\ v \neq 0 \text{ s.t. } \tilde{X}_- v \neq 0$$

- This means: there exists a coefficient vector v that generates a nonzero state

$$\tilde{x}_0 = \tilde{X}_- v = \Phi_-^X X_- v$$

- Then $\tilde{x}_0 \in \mathcal{V}^*(\tilde{\mathcal{D}}) \rightarrow$ Strong observability is destroyed.

Attacker's goal: To embed a nonzero state into $\mathcal{V}^*(\tilde{\mathcal{D}})$

Malicious Model and Trajectories

Assumption 2:

The attacker has a virtual model $A_{\text{mal}} \in \mathbb{R}^{n\times n}$ and its eigenpair $(\tilde{\lambda}, \tilde{x}_0)$ such that $(A_{\text{mal}}, B, C, D)$ is weakly unobservable and $A_{\text{mal}}\tilde{x}_0 = \tilde{\lambda}\tilde{x}_0$.

- The attacker constructs target vectors:

$$\tilde{x}_0 = \Phi_-^X X_- v \quad \Rightarrow \text{ To be embedded into } \mathcal{V}^*(\tilde{\mathcal{D}})$$

$$\tilde{u}_0 = \Phi_-^U U_- v$$

$$\tilde{x}_1 = \Phi_+^X X_+ v = \tilde{\lambda}\tilde{x}_0 + B\tilde{u}_0$$

$$\tilde{y}_0 = \Phi_-^Y Y_- v = D\tilde{u}_0$$

One-step virtual trajectories of the malicious model $(A_{\text{mal}}, B, C, D)$

v: injection direction in coefficient space

- Additional constraint: $C\tilde{x}_0 = 0$

How to Construct Attack Transformation

Lemma 2: $Z \in \{X_-, X_+, U_-, Y_-\}$

Given data $Z \in \mathbb{R}^{m_Z \times T}$, vector $v \in \mathbb{R}^T \setminus \{0\}$, and target $z_{\text{tar}} \in \mathbb{R}^{m_Z}$, for any $\xi_Z \in \mathbb{R}^{m_Z}$ with $\xi_Z^\mathsf{T} Z v = 1$ and $\xi_Z^\mathsf{T} z_{\text{tar}} \neq 0$, let

$$\Phi_Z := I_{m_Z} + (z_{\text{tar}} - Zv)\xi_Z^\mathsf{T}.$$

Then:

(i) Φ_Z is nonsingular

(ii) $\Phi_Z Z v = z_{\text{tar}}$ (injection)

(iii) For any $w \in \ker(\xi_Z^\mathsf{T} Z)$, $\Phi_Z Z w = Z w$ (invariance)

- Along v, the original trajectory replaced by the target: $Zv \to z_{\text{tar}}$
- Original unobservable trajectory subspace unchanged: $Zw \to Zw$

  → For stealthiness

When is the Attack Feasible?

Proposition 1:

The following are equivalent for each $Z \in \{X_-, X_+, U_-, Y_-\}$:

(i) Design feasibility: $\exists v \in \mathcal{J}^*(\mathcal{D})^\perp \setminus \{0\}$, ξ_Z with $\xi_Z^\mathsf{T} Z v = 1$ and $\xi_Z^\mathsf{T} Z w = 0, \forall w \in \mathcal{J}^*(\mathcal{D})$.

(ii) Geometric condition: $\exists v \in \mathcal{J}^*(\mathcal{D})^\perp \setminus \{0\}$ s.t. $Zv \notin Z\mathcal{J}^*(\mathcal{D})$.

(iii) Dimensional condition: $\dim Z\mathcal{J}^*(\mathcal{D}) < \operatorname{rank} Z$

Additional data direction is available for injecting a malicious trajectory.

Moreover, if any of (i)-(iii) holds, then the linear map $\Phi_a : \mathbb{D} \to \mathbb{D}$ exists and it is a solution to Problem 1.

- Condition (iii) is checkable from the original data.

Main Result 1: Subspace Expansion

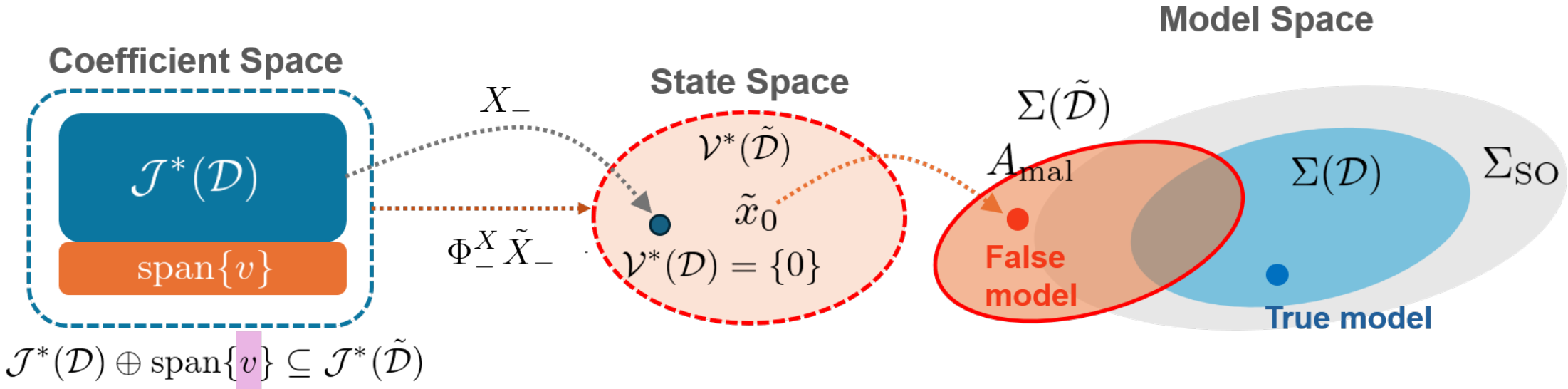
Theorem 1: Under the conditions of Proposition 1, after applying Algorithm 1:

(i) $\mathcal{J}^*(\mathcal{D}) \oplus \operatorname{span}\{v\} \subseteq \mathcal{J}^*(\tilde{\mathcal{D}})$.

(ii) If $\Sigma(\tilde{\mathcal{D}}) \neq \varnothing$, then $\Sigma(\tilde{\mathcal{D}}) \cap \{A_{\text{mal}} \in \mathbb{R}^{n\times n} \mid A_{\text{mal}}\tilde{x}_0 = \tilde{\lambda}\tilde{x}_0\} \neq \varnothing$.

(iii) $\tilde{\mathcal{D}}$ is not informative for strong observability.

- The attack expands the weakly unobservable subspace while preserving its original directions.



Minimum-Norm Attacks

- For stealthiness: Minimize detectability under statistical tests.
- Vulnerability metric: Quantify the smallest distortion needed to break informativity.
- Data perturbation: $\Delta_Z := \tilde{Z} - Z = (z_{\text{tar}} - Zv)\xi_Z^{\mathsf{T}} Z$
- Objective: Minimize $\|\Delta\mathcal{D}\|_{\text{F}}^2$

Main Result 2: Lower Bound on Attack Norm

Theorem 2:

Suppose there is no noise in original data $\mathcal{D}$. Then:

$$\inf_{\lambda \in \mathbb{C},\, v \in \mathcal{F}} \|\Delta \mathcal{D}\|_{\mathrm{F}} \geq d_{\mathrm{UNOBS}}(\Sigma(\mathcal{D}))\, \sigma_{\min}(X_{-})$$

- $d_{\mathrm{UNOBS}}(\Sigma(\mathcal{D})) := \inf_{A \in \Sigma(\mathcal{D})} d_{\mathrm{UNOBS}}(A)$: Distance of the model set from unobservability

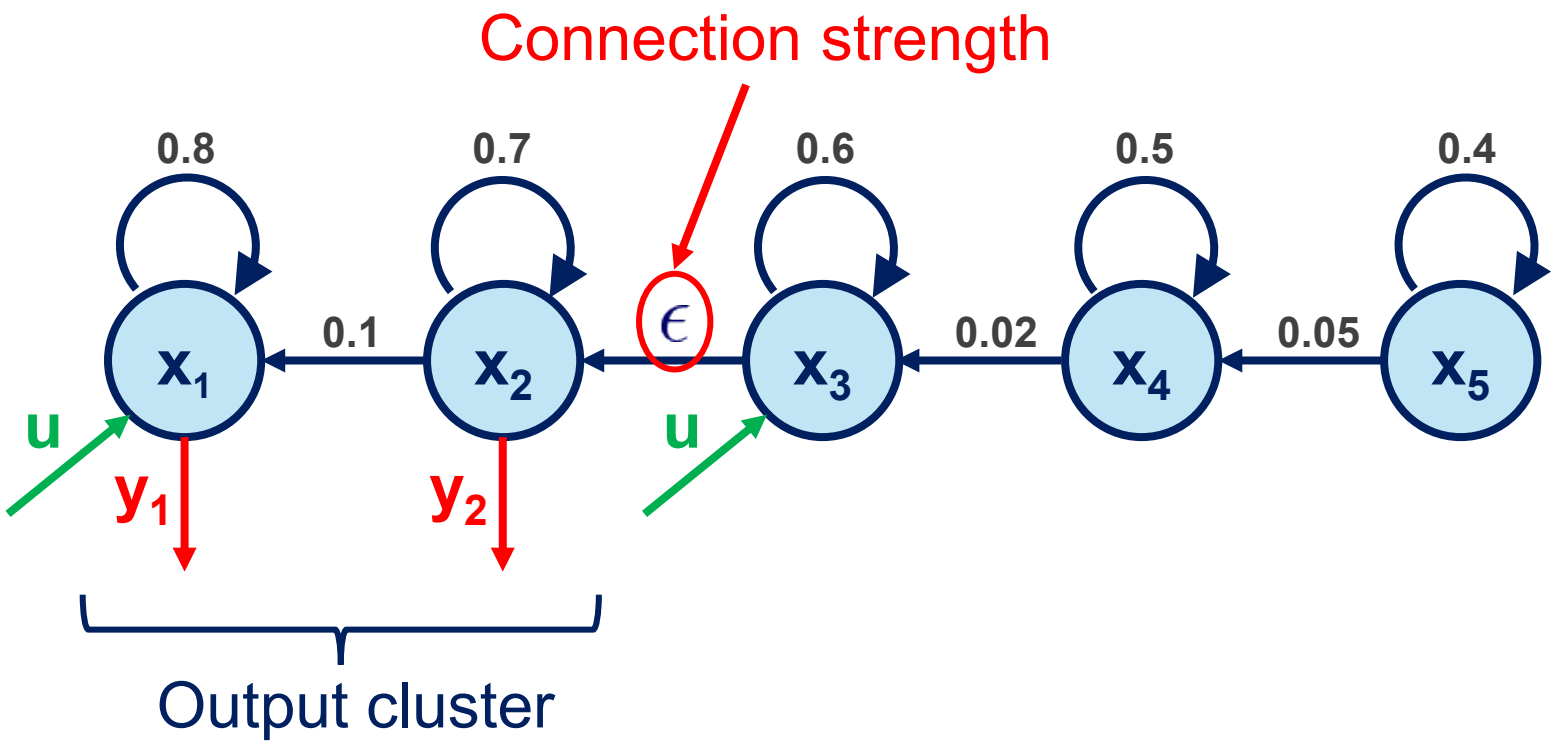
$$d_{\mathrm{UNOBS}}(A) := \inf_{\lambda \in \mathbb{C}} \sigma_{\min}\left(\begin{bmatrix} \lambda I_n - A \\ C \end{bmatrix} \right)$$

- $\sigma_{\min}(X_{-})$: Smallest singular value of the state data matrix

- Interpretation:
  - — Systems close to unobservability → vulnerable to small attacks
  - — Well-conditioned data (large $\sigma_{\min}(X_{-})$)→ larger perturbation required

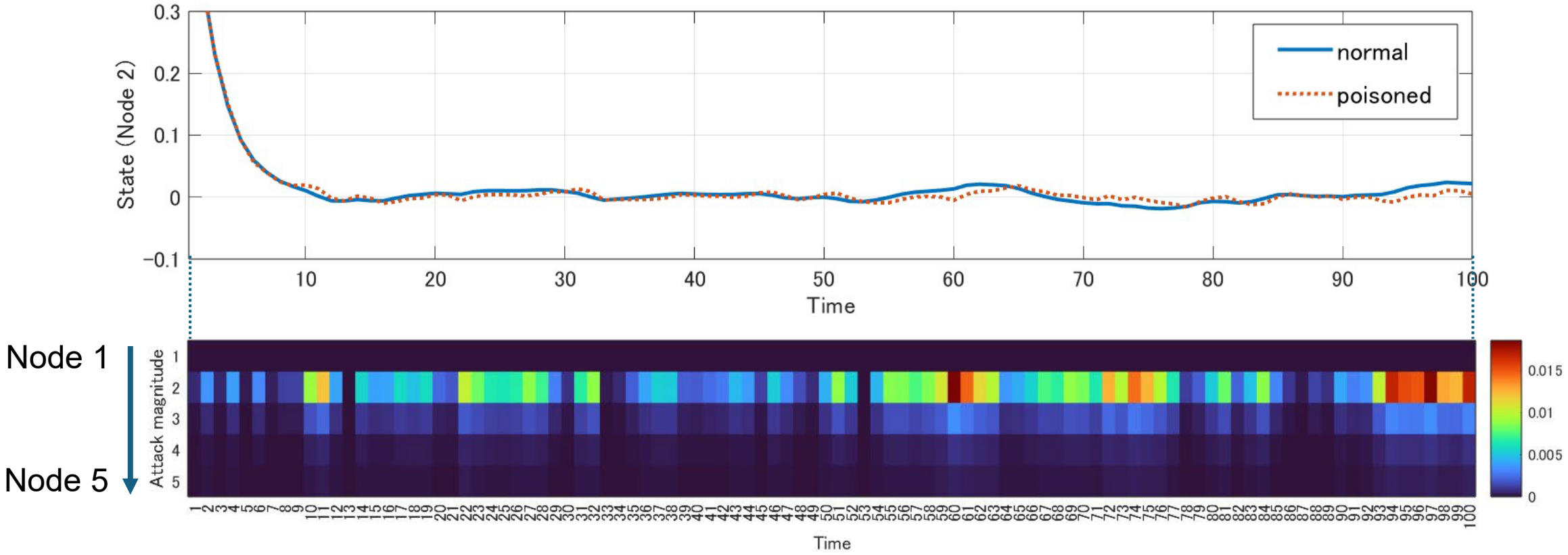
R. Eising (1984), *Syst. Control Lett.*

Linear Dynamical Network Example



- Unknown $A_{\text{true}} \Leftrightarrow$ Unknown network structure
- (A_{true}, C): Observable
- Time horizon T = 100, random initial states, no noise.

Attack Structure (ε=0.1)



Key observations

The attack highly localized near measured nodes:

Node 2 — 96.4%, Node 3— 3.3%, Others: negligible

⇒ **Protect data on vulnerable nodes based on network topology**

Observability vs. Minimum Attack Norm

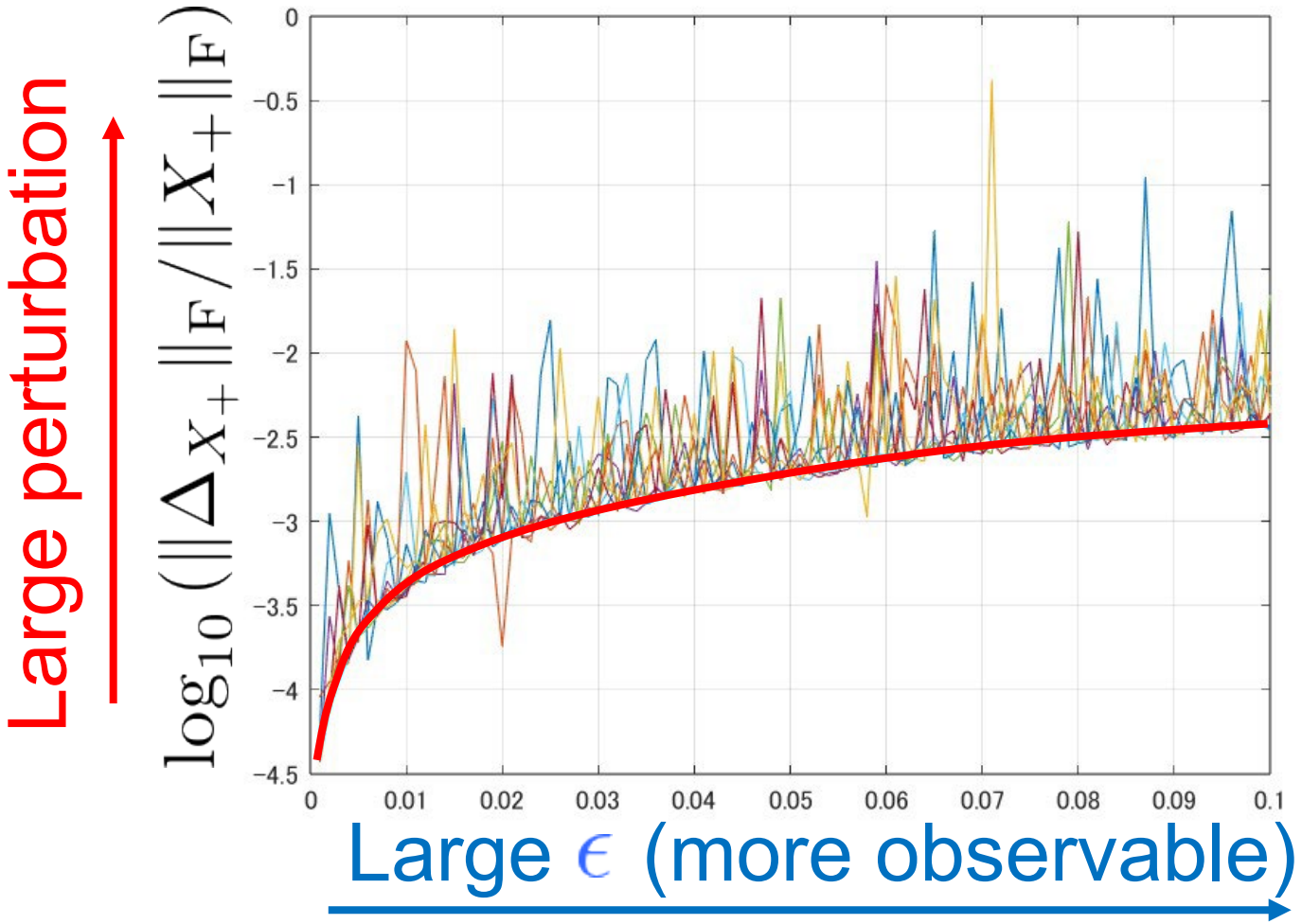
Theorem 2

$$\inf_{\lambda \in \mathbb{C},\, v \in \mathcal{F}} \|\Delta \mathcal{D}\|_{\mathrm{F}} \geq d_{\mathrm{UNOBS}}(\Sigma(\mathcal{D}))\, \sigma_{\min}(X_-)$$

- Monte-Carlo Simulation

  Only successful attacks plotted

- Smaller ε makes the system closer to be unobservable.



Conclusion

- Data-driven approach for networked control:

  Imperfect or malicious data can fundamentally affect data-driven control.

- Data poisoning can destroy observability informativity while preserving apparent data richness.

- Manipulating the observability structure may create new vulnerabilities to stealthy attacks exploiting zero dynamics.

Future work: Extend the framework to controllability / stabilizability, and develop attack detection and data-recovery methods.

→ **Toward reliable and secure data-driven control !**

References

- I. Takaki, A. Cetinkaya, and H. Ishii, Data poisoning attacks on informativity for observability: Invariance-based synthesis, *arXiv*, 2026.
- —, Trade-off in quantization between data-driven design and control inputs, *Proc. 10th IFAC Conference on Networked Systems (NecSys'25),* 2025.
- —, Data-driven quantized control of unknown linear systems with uncertainties, *Trans. SICE*, 2025 (in Japanese).