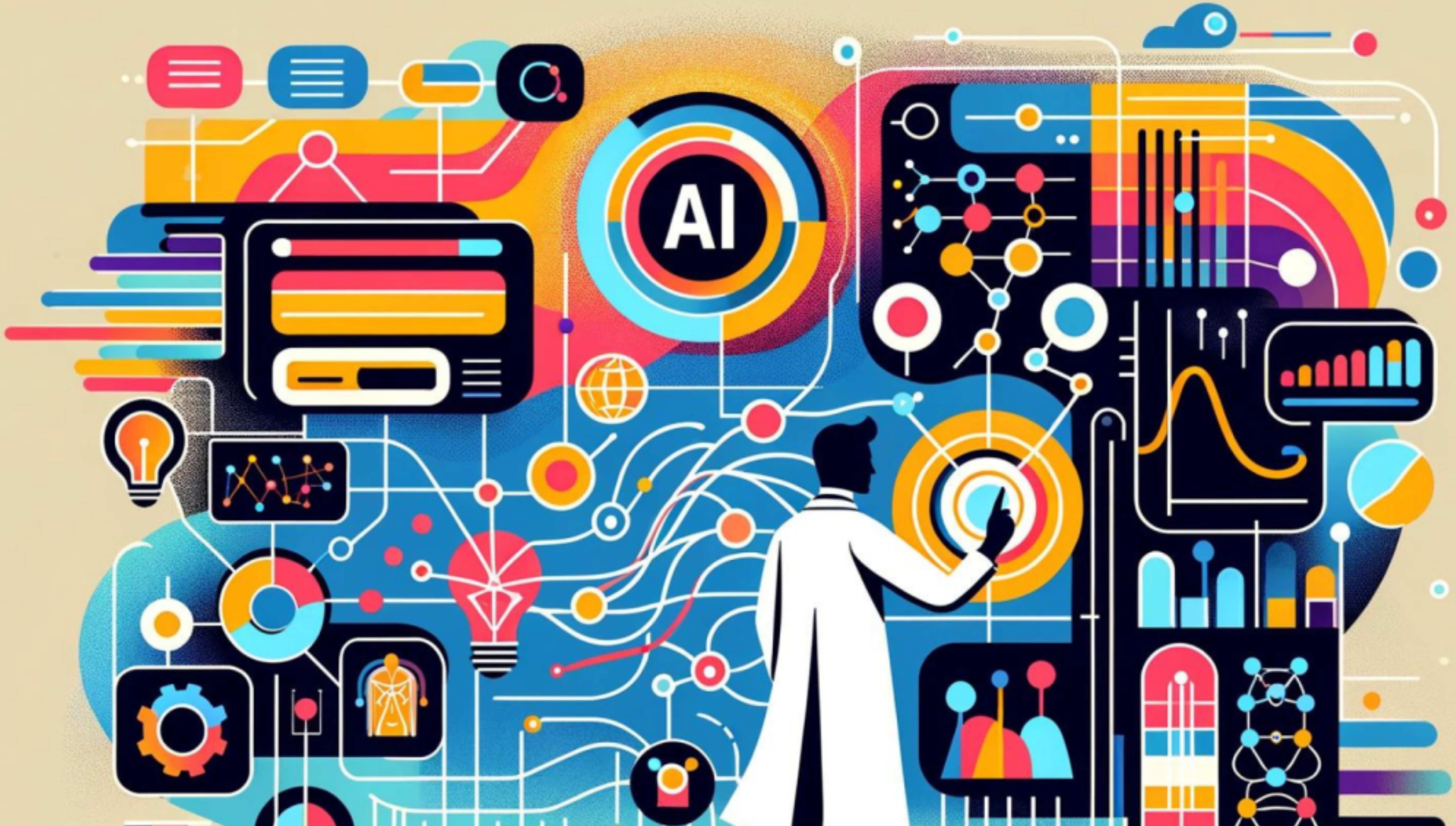
A Tutorial on the Non-Asymptotic Theory of System Identification

Ziemann, Tsiamis, Lee, Jedra, Matni & Pappas

AI for Science Seminar FS2024

Damiano Meier, Lino Hofstetter

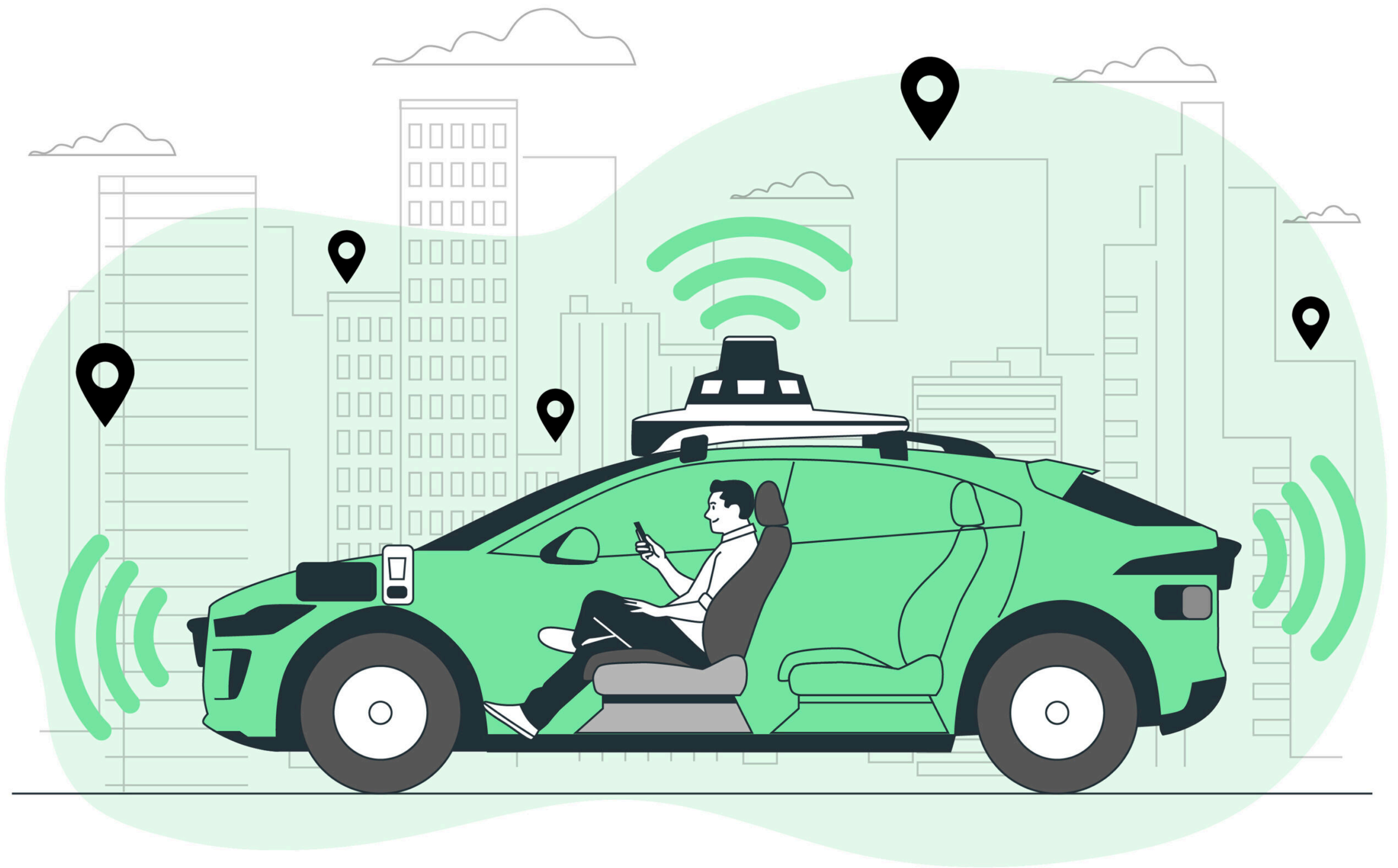
Machine Learning is everywhere…



… but it isn’t perfect!

System Identification

- **Energy Optimization**: Efficient energy use in buildings
- **Adaptive Suspensions**: Vehicles adjusting to road conditions
- **Financial Markets**: Adapting to market fluctuations



- **Instable Energy Grid, Car Crashes & Financial Loss**

The Paper

arXiv > eess > arXiv:2309.03873

Electrical Engineering and Systems Science > Systems and Control

[Submitted on 7 Sep 2023]

A Tutorial on the Non-Asymptotic Theory of System Identification

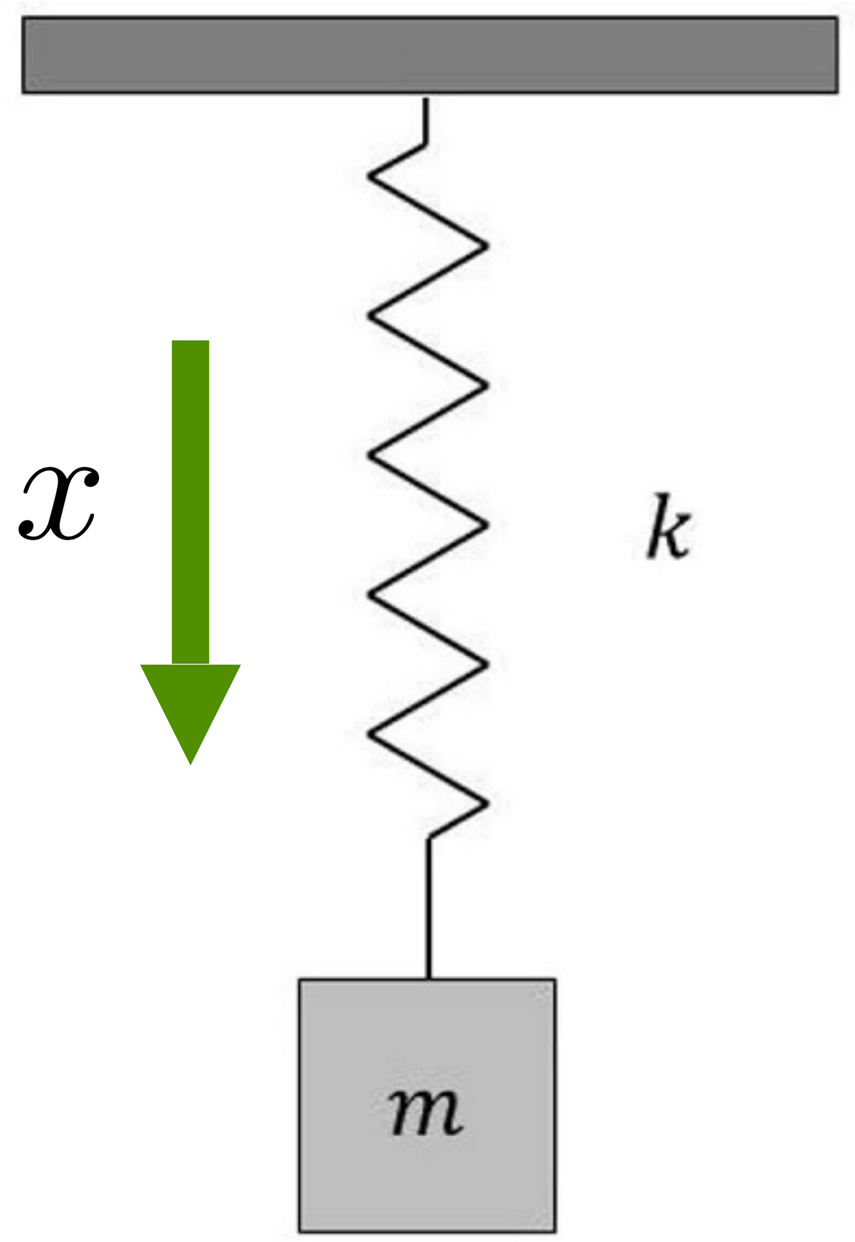
Ingvar Ziemann, Anastasios Tsiamis, Bruce Lee, Yassir Jedra, Nikolai Matni, George J. Pappas

Outline

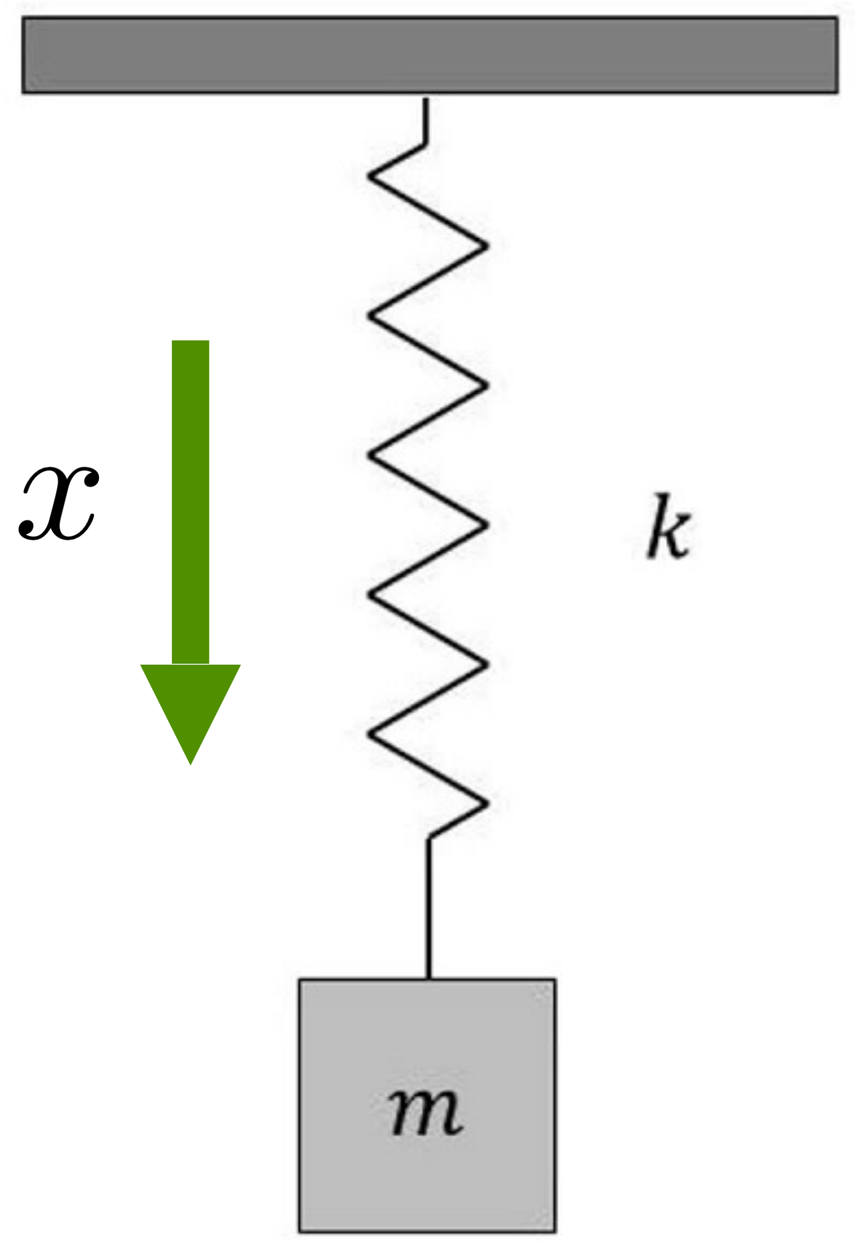
- **Introduction to System Identification**
- **Main Result of the Paper**
- **Proof Outline**
- **First Step of the Proof in Detail**
- **Extending the Results**
- **Discussion**: impact of the paper
- **Conclusion**: our personal opinion
- **Questions**

Introduction to System Identification

Intro to System Identification

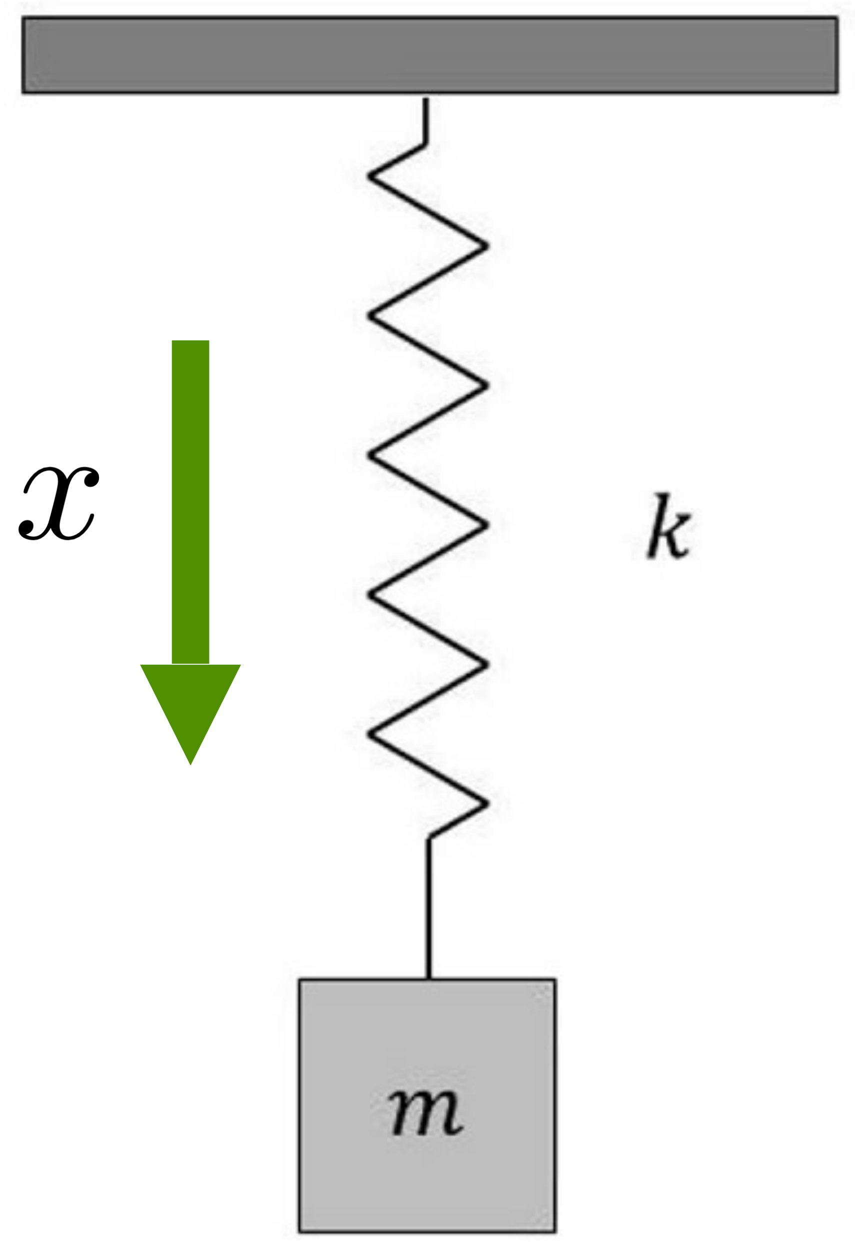


Intro to System Identification



Model: $mg = kx$

Intro to System Identification

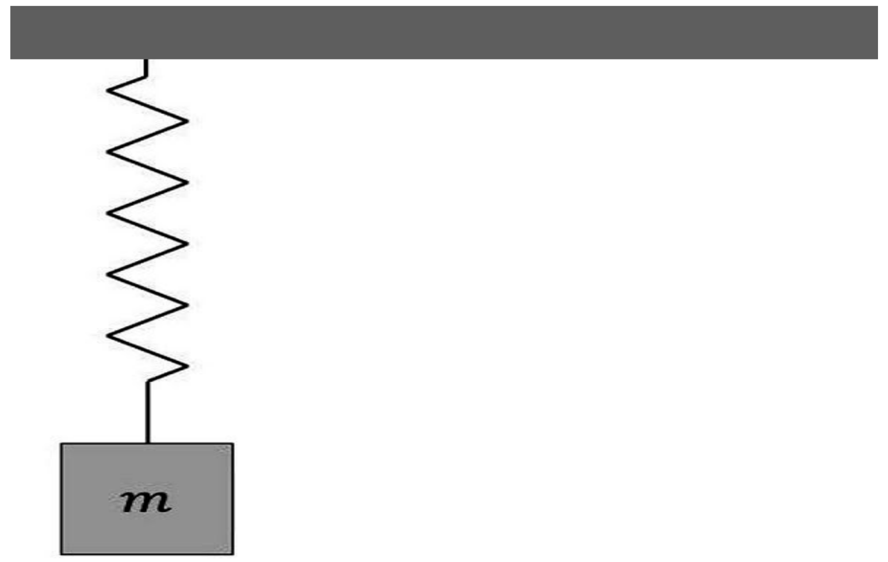


$$\text{Model: } mg = kx$$

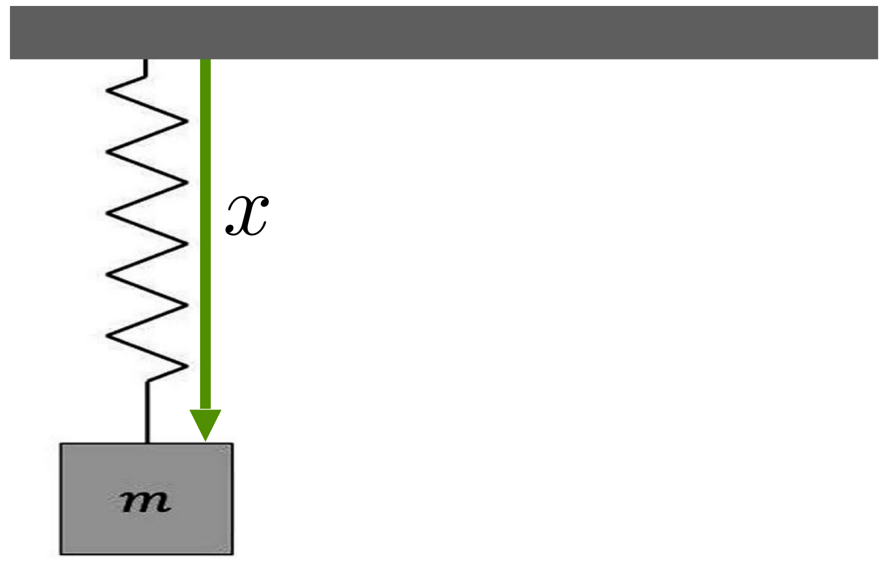
White Box Approach: Understanding of dynamics of the system

Intro to System Identification

Black Box Approach

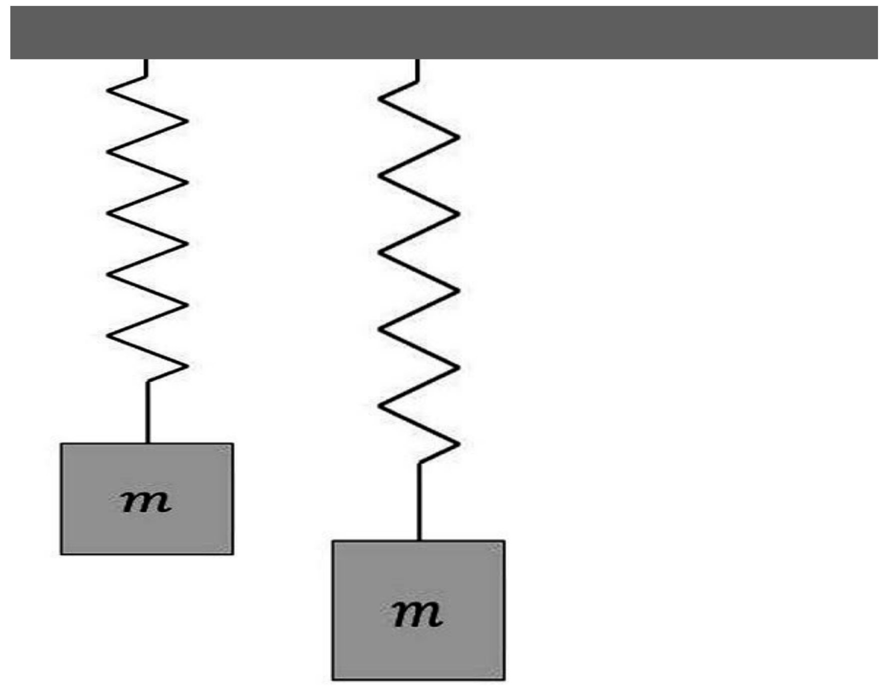


→ System with **unknown dynamics** →

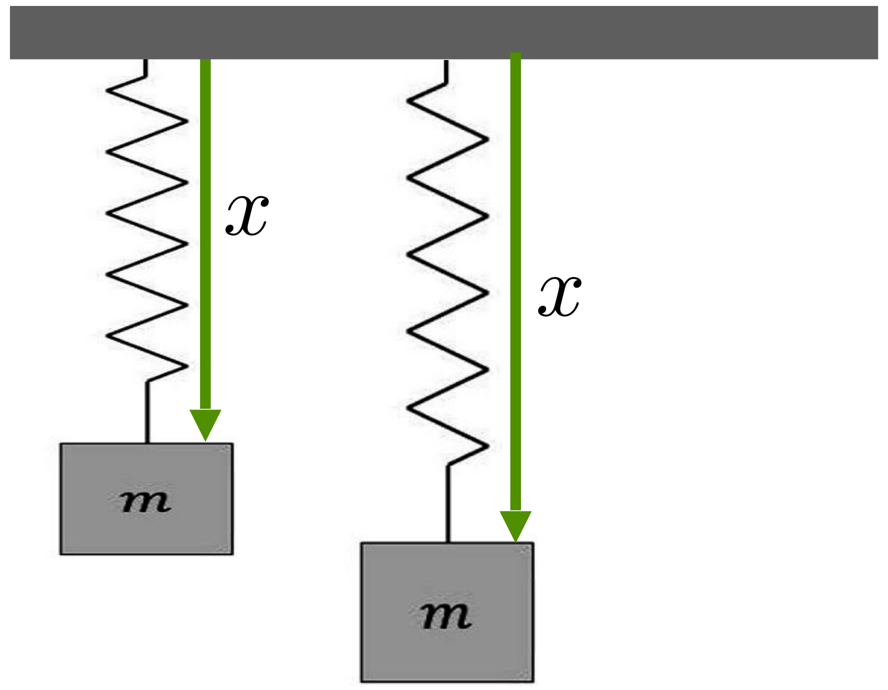


Intro to System Identification

Black Box Approach

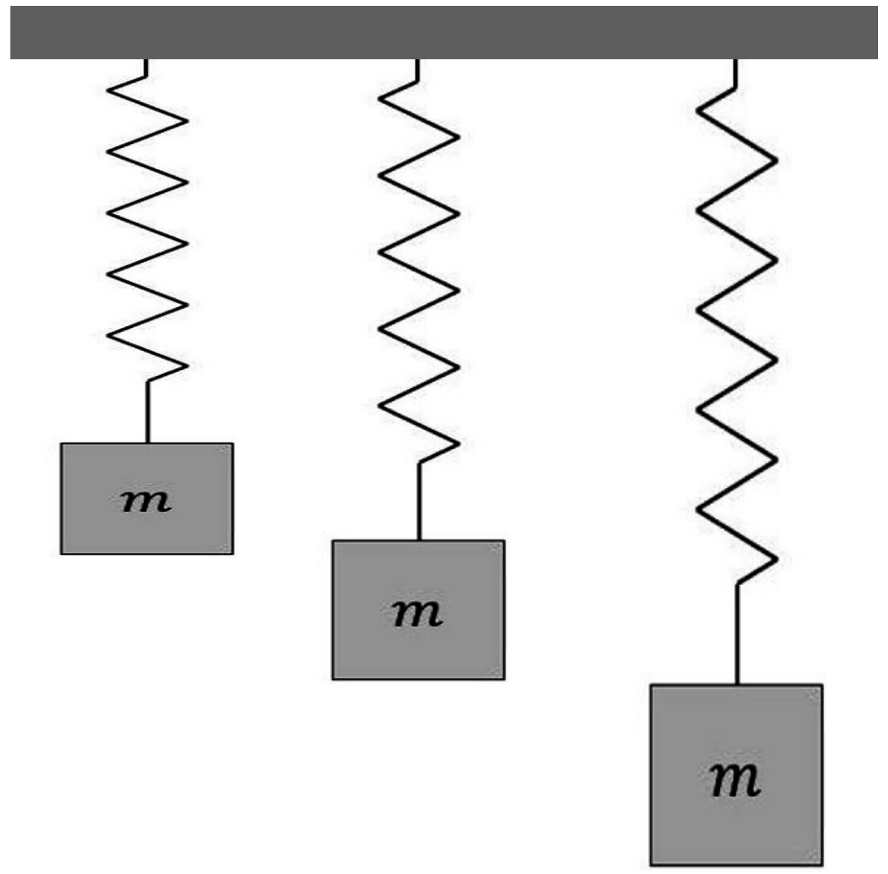


System with **unknown dynamics**

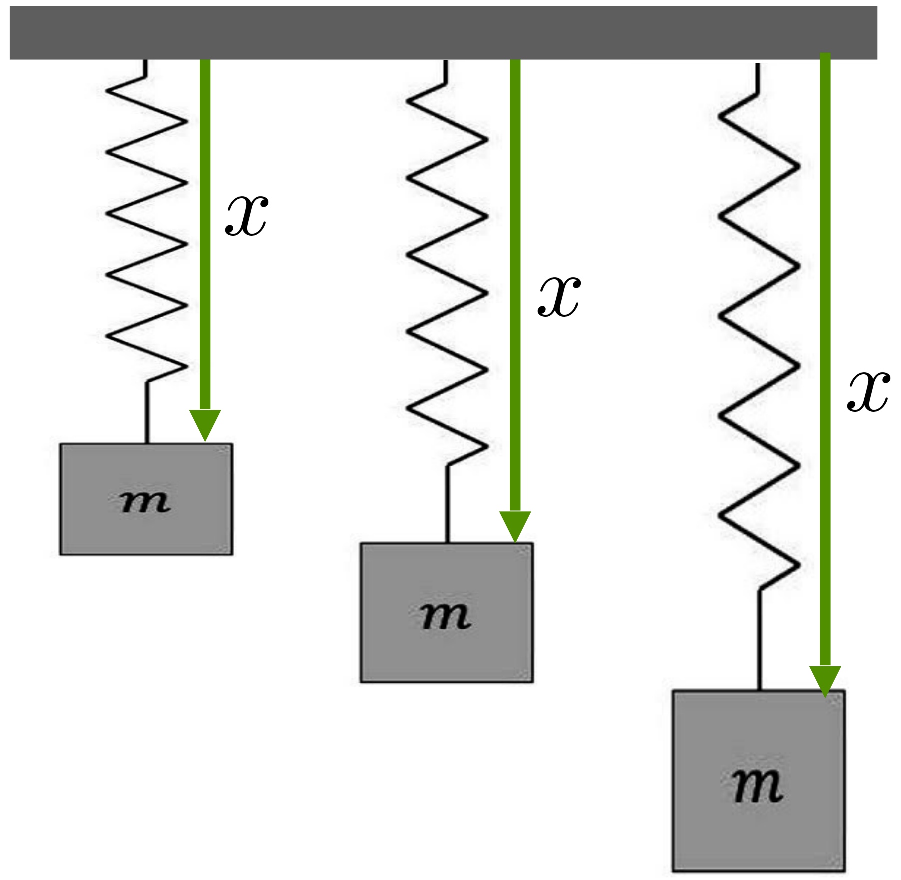


Intro to System Identification

Black Box Approach

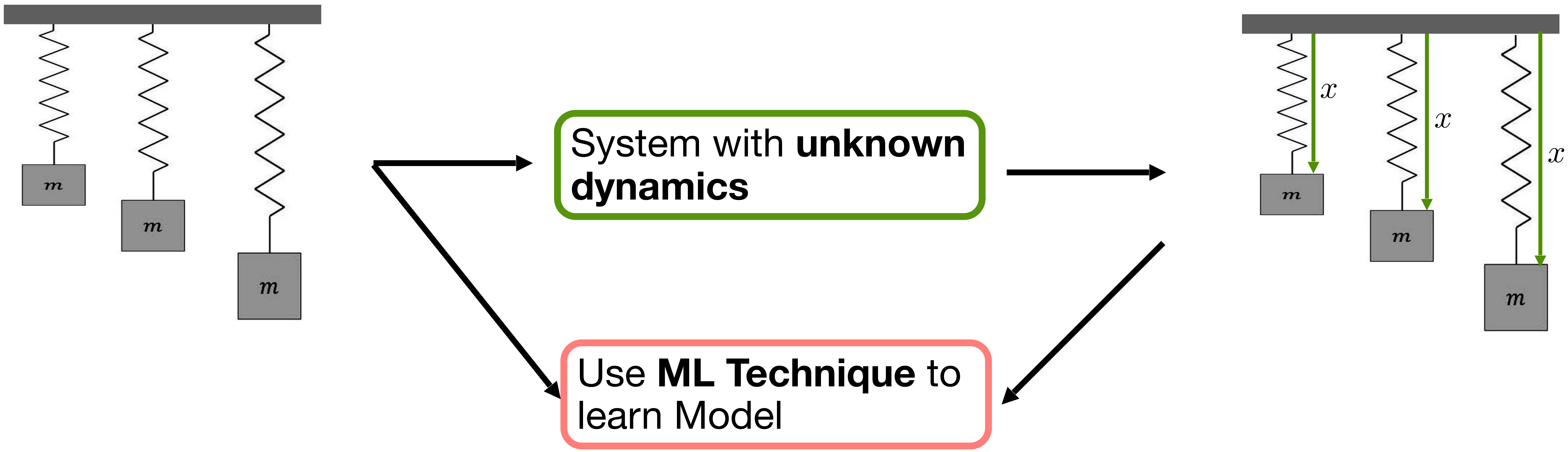


System with **unknown dynamics**



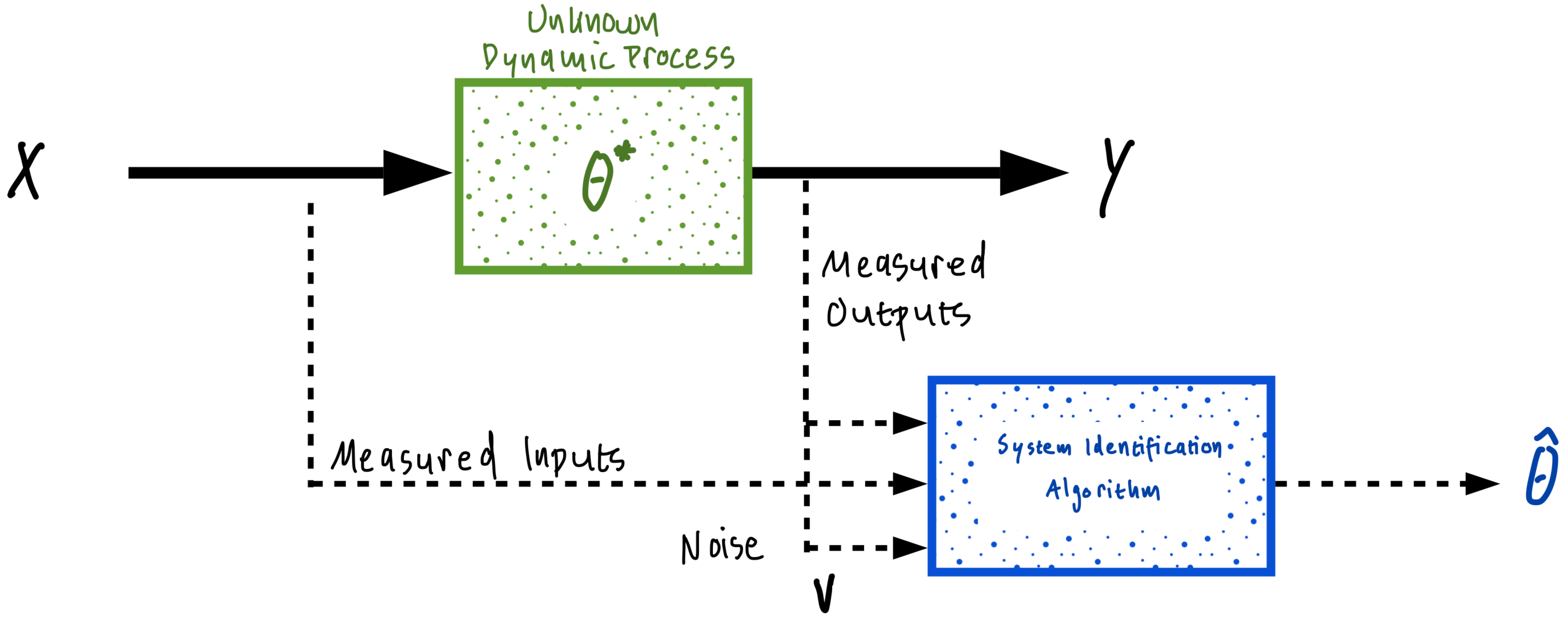
Intro to System Identification

Black Box Approach



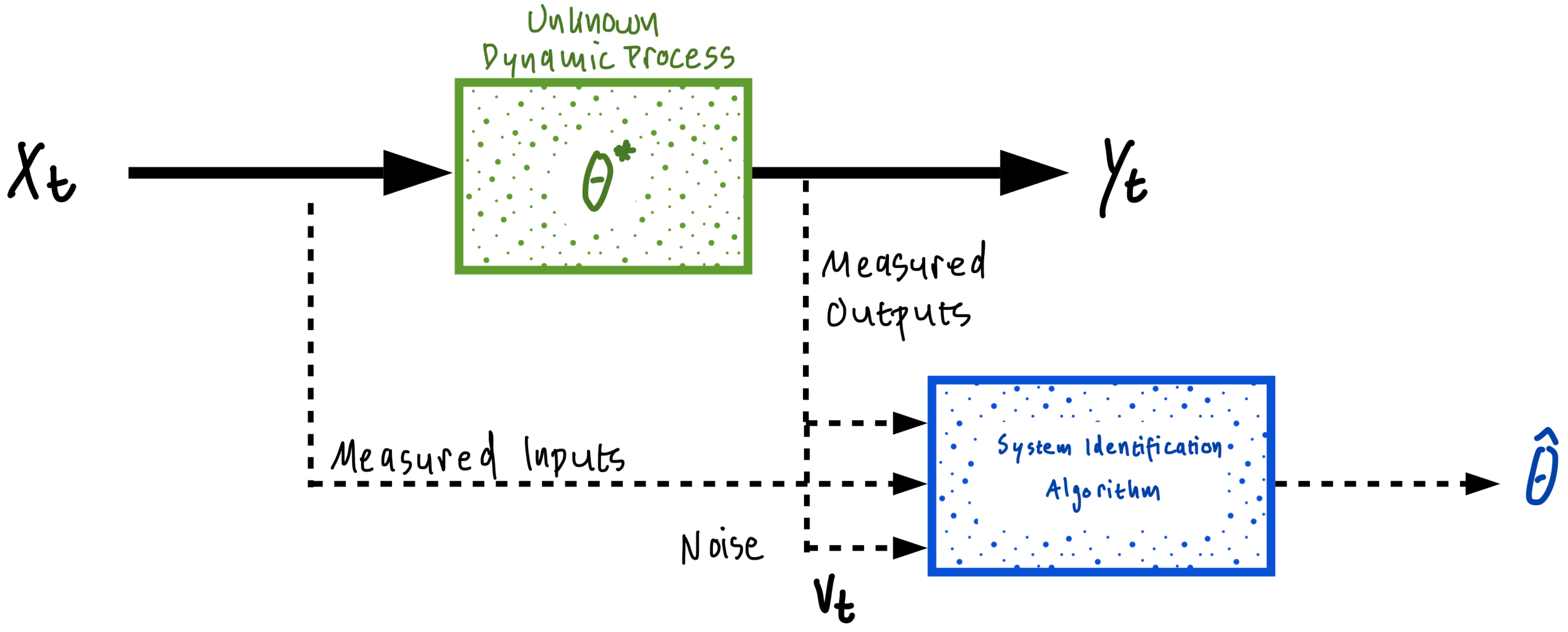
What is a linear time-series model?

$$Y = \theta^\star X + V$$



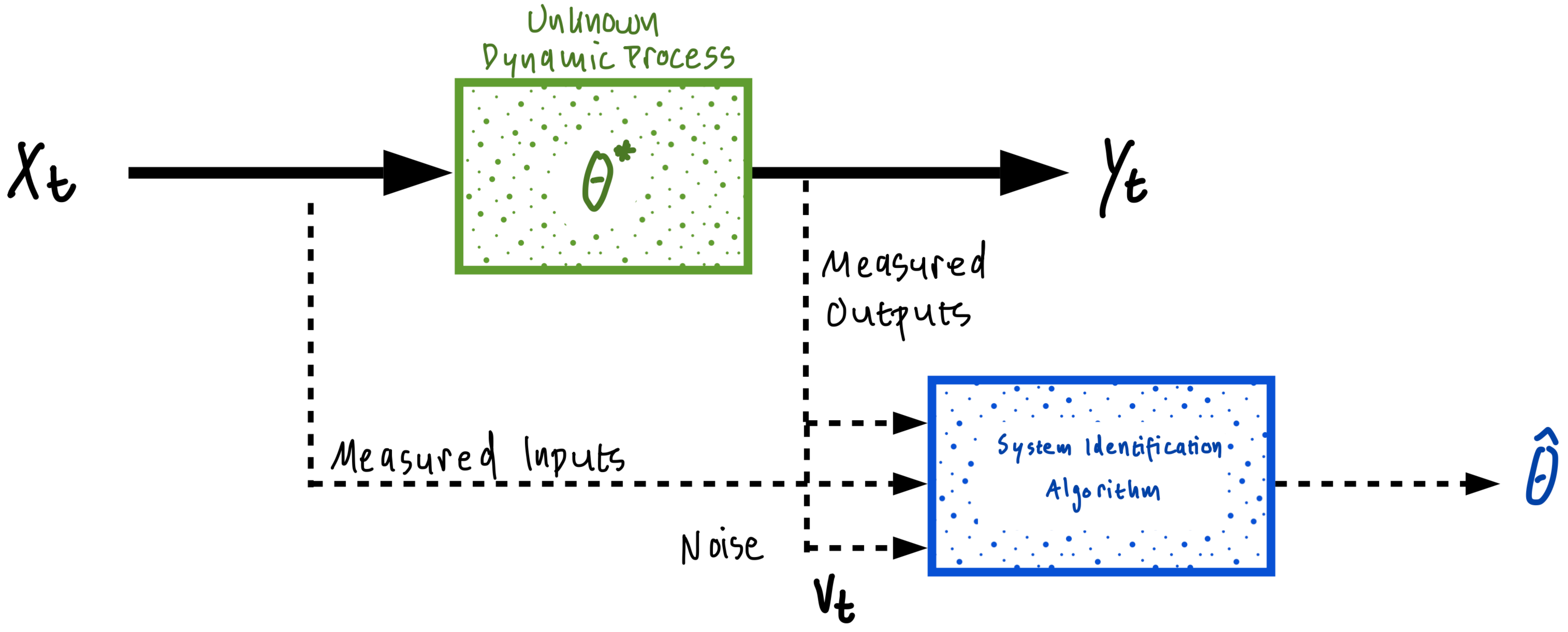
What is a linear time-series model?

$$Y_t = \theta^\star X_t + V_t \quad t = 1, 2, \ldots, T$$



What is a linear time-series model?

$$Y_t = \theta^\star X_t + V_t \quad t = 1, 2, \ldots, T$$



Autoregressive Exogenous Models

$$Y_t = \sum_{i=1}^{p} A_i^\star Y_{t-i} + \sum_{j=1}^{q} B_i^\star U_{t-j} + W_t$$

$$\begin{aligned} Y_t \in \mathbb{R}^{d_U} &:= \text{System outputs at time } t. \\ A_i^*, B_i^* &:= \text{Unkown ARX parameters} \\ U_j &:= \text{User specified input at step } j. \\ W_t &:= \text{Noise term at time } t. \end{aligned}$$

ARX model as a linear system

ARX Model

$$X_t = \begin{bmatrix} Y_{t-1:t-p}^{\top} & U_{t-1:t-q}^{\top} \end{bmatrix}^{\top}; \quad \theta^\star = \begin{bmatrix} A_{1:p}^\star & B_{1:q}^\star \end{bmatrix}; \quad V_t = W_t$$

$$Y_t = \theta^\star X_t + V_t \quad t = 1, 2, \ldots, T$$

Main Result

Main Result

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions* V.1, V.2 *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

Main Result

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions* V.1, V.2 *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((p d_{\mathsf{Y}} + q d_{\mathsf{U}}) \log \frac{p d_{\mathsf{Y}} + q d_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

Quality of Approximation

θ^* := True System Parameters

$\hat{\theta}$:= Estimated Parameters (LSE)

This term explains how close the approximated parameters are to the true parameters

Main Result

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions* V.1, V.2 *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

PREREQUISITES

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

Quality of Approximation

$\theta^* :=$ True System Parameters

$\hat{\theta} :=$ Estimated Parameters (LSE)

This term explains how close the approximated parameters are to the true parameters

Main result: says how close the approximation is to the real system for some prerequisites

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions* V.1, V.2 *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((p d_{\mathsf{Y}} + q d_{\mathsf{U}}) \log \frac{p d_{\mathsf{Y}} + q d_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

ARX System Parameters

$$p, q, d_y, d_U$$

$$Y_t = \sum_{i=1}^{p} A_i^\star Y_{t-i} + \sum_{j=1}^{q} B_i^\star U_{t-j} + W_t$$

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under* *Assumptions V.1, V.2* *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

ARX System Parameters

$$p, q, d_y, d_U$$

System Assumptions

V.1 Non-explosive system

V.2 Noise is white noise

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions V.1, V.2 for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

ARX System Parameters

p, q, d_y, d_U

System Assumptions

V.1 Non-explosive system

V.2 Noise is white noise

Approximation Quality

Probability that bound holds directly affects the size of the bound

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let $(Y_{1:T}, U_{0:T-1})$ be single trajectory input-output samples generated by system (38) under Assumptions V.1, V.2 for some horizon T. Fix a failure probability $0 < \delta < 1$ and a time index $\tau \geq \max\{p, q\}$. Let $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, where T_0 is defined in (46). If $T \geq T_{\mathsf{pe}}(\delta, \tau)$, then with probability at least $1 - \delta$*

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log \det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C is a universal constant, i.e., it is independent of system, confidence δ and index τ.

ARX System Parameters

p, q, d_y, d_U

System Assumptions

V.1 Non-explosive system

V.2 Noise is white noise

Approximation Quality

Probability that bound holds directly affects the size of the bound

Required Iterations

For given parameters, computes after how many time step the bound holds

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions V.1, V.2 for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \boxed{\frac{C}{\mathsf{SNR}_\tau T}} \Big((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right) \Big), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

Signal to Noise Ratio

High SNR means good data with small noise

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions V.1, V.2 for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

Signal to Noise Ratio

High SNR means good data with small noise

Input-Output Samples

Main Result - Prerequisites

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions* V.1, V.2 *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\Big((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\Big), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

Signal to Noise Ratio

High SNR means good data with small noise

Input-Output Samples

Noise Covariance Matrices

additional constraint determined by noise in the system

Main Result

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions* V.1, V.2 *for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then with probability at least* $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}}) \log \frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T \Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

How can we prove this?

Proof Outline

$$Y_t = \theta^\star X_t + V_t$$

$$\theta^\star = \begin{bmatrix} A^\star_{1:p} & B^\star_{1:q} \end{bmatrix}$$

$$X_t = \begin{bmatrix} Y^\top_{t-1:t-p} & U^\top_{t-1:t-q} \end{bmatrix}^\top$$

1. ARX Model as Linear System

Proof Outline

$$\widehat{\theta} \in \underset{\theta \in \mathrm{M}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \|Y_t - \theta X_t\|_2^2$$

1. ARX Model as Linear System

2. Fix size of system θ and define the Least Squares Estimator $M \in \mathbb{R}^{dy \times dx}$

Proof Outline

$$\widehat{\theta} \in \operatorname*{argmin}_{\theta \in \mathrm{M}} \frac{1}{T} \sum_{t=1}^{T} \|Y_t - \theta X_t\|_2^2$$

$$\widehat{\theta} = \left(\sum_{t=1}^{T} Y_t X_t^{\top} \right) \left(\sum_{t=1}^{T} X_t X_t^{\top} \right)^{\dagger}$$

1. ARX Model as Linear System
2. Fix size of system θ and define the Least Squares Estimator $M \in \mathbb{R}^{dy \times dx}$
3. **Compute the** (existing) **closed form solution**

Proof Outline

$$\widehat{\theta} - \theta^\star$$

$$= \left[\left(\sum_{t=1}^{T} V_t X_t^\intercal \right) \left(\sum_{t=1}^{T} X_t X_t^\intercal \right)^{-1/2} \right] \left(\sum_{t=1}^{T} X_t X_t^\intercal \right)^{-1/2}$$

1. ARX Model as Linear System
2. Fix size of system θ and define the Least Squares Estimator $M \in \mathbb{R}^{dy \times dx}$
3. Compute the (existing) closed form solution
4. **Find an expression for the precision of the estimation**

Proof Outline

$$\widehat{\theta} - \theta^\star$$

$$= \left[\left(\sum_{t=1}^{T} V_t X_t^\top \right) \left(\sum_{t=1}^{T} X_t X_t^\top \right)^{-1/2} \right] \left(\sum_{t=1}^{T} X_t X_t^\top \right)^{-1/2}$$

1. ARX Model as Linear System
2. Fix size of system θ and define the Least Squares Estimator $M \in \mathbb{R}^{dy \times dx}$
3. Compute the (existing) closed form solution
4. **Find an expression for the precision of the estimation**

Proof Outline

$$\widehat{\theta} - \theta^\star$$
$$= \left[\left(\sum_{t=1}^{T} V_t X_t^\top \right) \left(\sum_{t=1}^{T} X_t X_t^\top \right)^{-1/2} \right] \left(\sum_{t=1}^{T} X_t X_t^\top \right)^{-1/2}$$

1. ARX Model as Linear System
2. Fix size of system θ and define the Least Squares Estimator $M \in \mathbb{R}^{dy \times dx}$
3. Compute the (existing) closed form solution
4. Find an expression for the precision of the estimation
5. **Analyse the expression to find a bound on the estimation quality**

Proof Outline

This is a Matrix!

$$\widehat{\theta} - \theta^\star$$

$$= \left[\left(\sum_{t=1}^{T} V_t X_t^\intercal\right)\left(\sum_{t=1}^{T} X_t X_t^\intercal\right)^{-1/2}\right]\left(\sum_{t=1}^{T} X_t X_t^\intercal\right)^{-1/2}$$

Time-Scale Invariant

Controls growth

1. ARX Model as Linear System
2. Fix size of system θ and define the Least Squares Estimator $M \in \mathbb{R}^{dy \times dx}$
3. Compute the (existing) closed form solution
4. Find an expression for the precision of the estimation
5. Analyse the expression to find a bound on the estimation quality

The trick is to analyse left & right term separately!

Structure of the Paper

$$\widehat{\theta} - \theta^\star$$

$$= \left[\left(\sum_{t=1}^{T} V_t X_t^\mathsf{T} \right) \left(\sum_{t=1}^{T} X_t X_t^\mathsf{T} \right)^{-1/2} \right] \left(\sum_{t=1}^{T} X_t X_t^\mathsf{T} \right)^{-1/2}$$

Time-Scale Invariant

Controls growth

Section I: Introduction to topic

Section II: Math Prerequisites

Section III: Analyse **right term**

Section IV: Analyse **left term**

Section V: Full proof

Section VI - **VII**: Extending results

First Step of Proof in Detail

Proof in Detail

A lot of Prerequisites

↓

Theorem V.1: After certain amount of steps, the approximation is smaller than the given bound with given probability

↓

Prove the theorem under the given assumptions

Proof in Detail

A lot of Prerequisites

These Prerequisites aren't arbitrary but very important to be able to prove the Theorem!

Theorem V.1: After certain amount of steps, the approximation is smaller than the given bound with given probability

How should we choose those prerequisites?

Prove the theorem under the given assumptions

This is the first part of the proof!

Proof in Detail

How the **measured data** generated by the system looks, is very important

Not enough **excitation** → data doesn't capture all possible system behaviour → **lower bound** the excitation

Proof in Detail

How the **measured data** generated by the system looks, is very important

Not enough **excitation** → data doesn't capture all possible system behaviour → **lower bound** the excitation

Too much **noise** → **overfitting** to high noise levels → **upper bound** the noise

Proof in Detail

How the **measured data** generated by the system looks, is very important

Not enough **excitation** → data doesn't capture all possible system behaviour → **lower bound** the excitation

Too much **noise** → **overfitting** to high noise levels → **upper bound** the noise

bad values → numerical **instability** → ensure **non-explosivity**

Proof in Detail

The constraints can now be defined mathematically

lower bound the excitation → $$\mathcal{E}_1 \triangleq \left\{ \widehat{\Sigma}_T \succeq \frac{1}{16} \Sigma_\tau \right\}$$

upper bound the noise → $$\mathcal{E}_2 \triangleq \left\{ \widehat{\Sigma}_T \preceq 3 \frac{p d_{\mathsf{Y}} + q d_{\mathsf{U}}}{\delta} \Sigma_T \right\}$$

ensure **non-explosivity** → $$\mathcal{E}_3 \triangleq \left\{ \left\| \sum_{t=1}^{T} W_t X_t^\top \left(\Sigma + T \widehat{\Sigma}_T \right)^{-1/2} \right\|_{\mathsf{op}}^2 \leq 4K^2 \log \left(\frac{\det \left(\Sigma + T \widehat{\Sigma}_T \right)}{\det(\Sigma)} \right) + 8 d_{\mathsf{Y}} K^2 \log 5 + 8K^2 \log \frac{3}{\delta} \right\}$$

Proof in Detail

The constraints can now be defined mathematically

lower bound the excitation → $\mathcal{E}_1$

upper bound the noise → $\mathcal{E}_2$

ensure **non-explosivity** → $\mathcal{E}_3$

$$\mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \geq 1 - \delta$$

→ **What can we do with this ?**

This is proven by combining various lemmas from the paper

Proof in Detail

Theorem V.1 (ARX Finite-Sample Bound). *Let* $(Y_{1:T}, U_{0:T-1})$ *be single trajectory input-output samples generated by system* (38) *under Assumptions V.1, V.2 for some horizon* T. *Fix a failure probability* $0 < \delta < 1$ *and a time index* $\tau \geq \max\{p, q\}$. *Let* $T_{\mathsf{pe}}(\delta, \tau) \triangleq \min\{t : t \geq T_0(t, \delta/3, \tau)\}$, *where* T_0 *is defined in* (46). *If* $T \geq T_{\mathsf{pe}}(\delta, \tau)$, *then* ***with probability at least*** $1 - \delta$

$$\|\widehat{\theta}_T - \theta^\star\|_{\mathsf{op}}^2 \leq \frac{C}{\mathsf{SNR}_\tau T}\left((pd_{\mathsf{Y}} + qd_{\mathsf{U}})\log\frac{pd_{\mathsf{Y}} + qd_{\mathsf{U}}}{\delta} + \log\det\left(\Sigma_T\Sigma_\tau^{-1}\right)\right), \quad (44)$$

where C *is a universal constant, i.e., it is independent of system, confidence* δ *and index* τ.

This bound also holds with probability at least $1 - \delta$

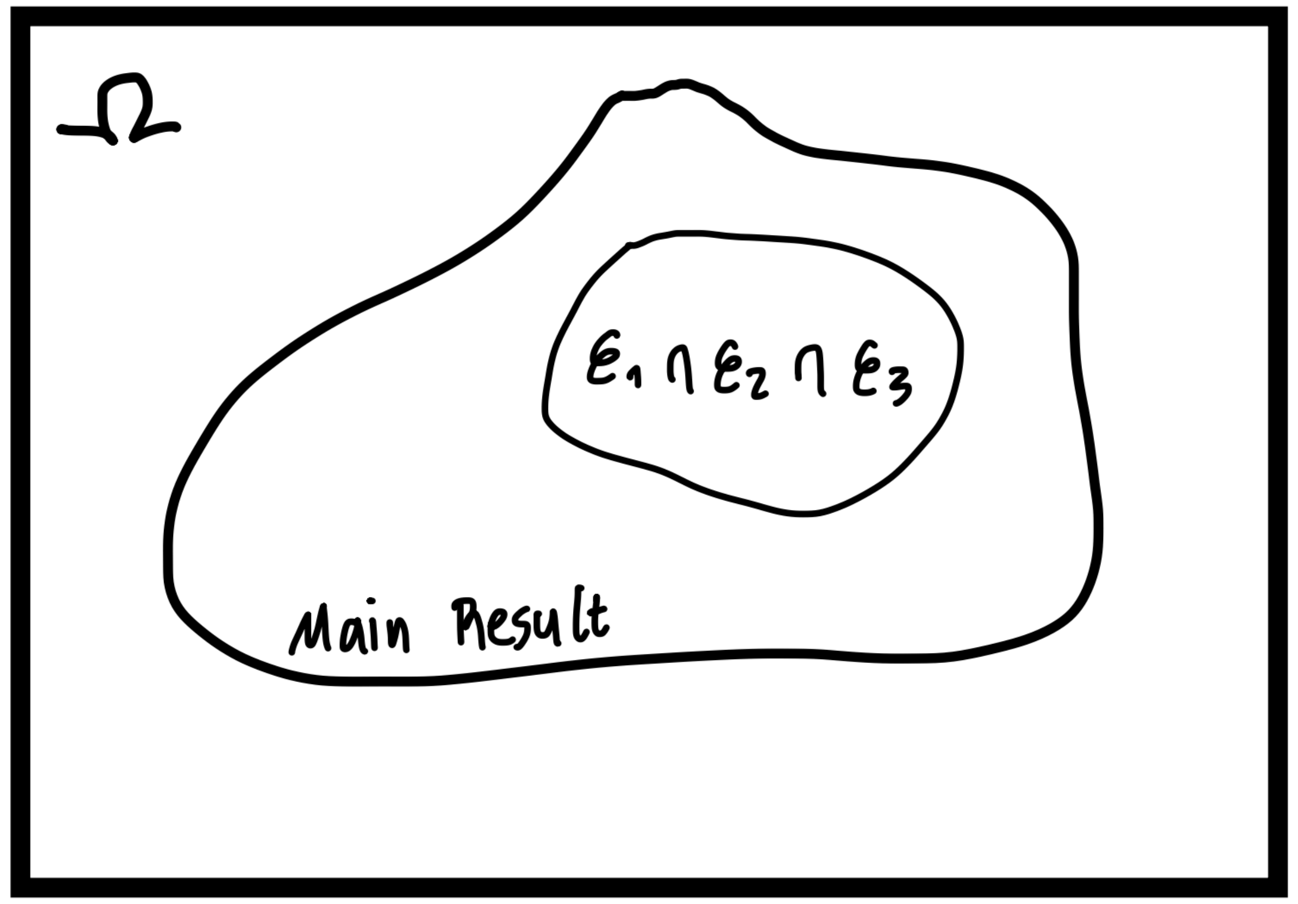
Proof in Detail

We can show that if the union of the three events holds, the main result holds

$$\mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \geq 1 - \delta$$

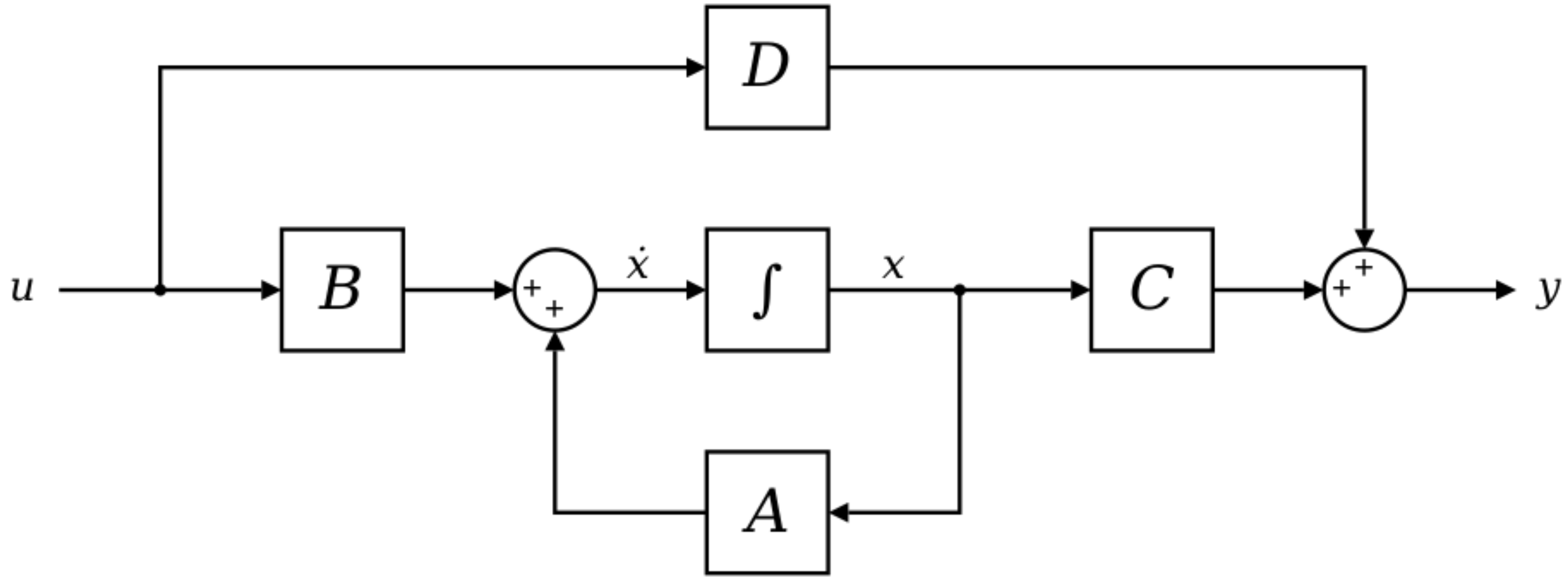
→ Show that if the three events hold, the main result holds, by using previous results ■

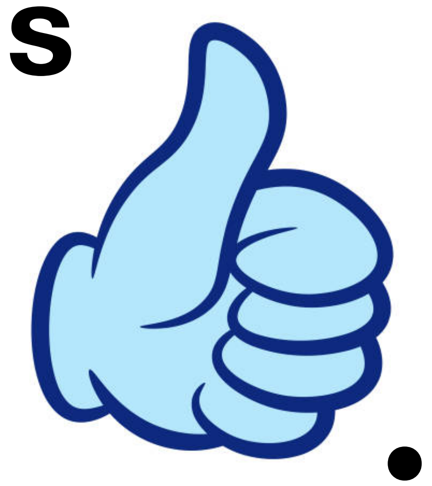
↑ This is proven by combining various lemmas from the paper



Extending the Result

- Theorem be extended to **state space models**





- **Macroeconomic Modeling**
- **Control Systems**
- **Climate Modeling**

Extending the Result

- Theorem be extended to **state space models**
- Paper also derives result for **less constraints** on matrix θ
- Presents ideas for extending results to **non-linear systems**

Discussion

- **Big contribution** to machine learning for control theory
- The proven bound on the approximation ratio is nearly optimal
- The constraints are realistic
- Therefore the LSE based approach for linear system identification can be used and it's performance is now well understood

- However it still remains open how good LSE is for non-linear systems
- Most systems are non-linear

Conclusion - our point of view

- The paper was very complicated
- Concepts from high-dimensional statistics we haven’t seen before
- Logical Flow of arguments wasn’t clear when reading at first

- Could motivate more why certain lemmas were introduced
- Online document with full proofs was very helpful

Questions?