# Parameter Estimation of Linear Dynamical Systems with Gaussian Noise

Léo Simpson

February 17, 2023

# Overview

1. Problem Statement
2. Description of the method
3. Open questions
4. The important particular case of linear time invariant systems (optional)

# General introduction

- This Ph.D. is part of the ELO-X program, an E.U. grant for Ph.D. programs in control coupled with learning.

# General introduction

- This Ph.D. is part of the ELO-X program, an E.U. grant for Ph.D. programs in control coupled with learning.

- The project is hosted by Tool-Temp: a company in Switzerland manufacturing Temperature Control Units.

# General introduction

- This Ph.D. is part of the ELO-X program, an E.U. grant for Ph.D. programs in control coupled with learning.
- The project is hosted by Tool-Temp: a company in Switzerland manufacturing Temperature Control Units.

# General introduction

- This Ph.D. is part of the ELO-X program, an E.U. grant for Ph.D. programs in control coupled with learning.
- The project is hosted by Tool-Temp: a company in Switzerland manufacturing Temperature Control Units.
- This project led to the submission of a paper for ECC 2023. [1]

[1] L. Simpson, A. Ghezzi, J. Asprion and M. Diehl,
"Parameter Estimation of Linear Dynamical Systems with Gaussian Noise,"
*arXiv preprint arXiv:2211.12302*, 2022.

# Parametric model

## Parametric Linear Dynamical Model with Gaussian Noise

Dynamical model:

$$\begin{aligned} x_{k+1} &= A_k(\alpha)x_k + b_k(\alpha) + w_k, && k = 0, \dots, N-1, \\ y_k &= C_k(\alpha)x_k + v_k, && k = 0, \dots, N, \end{aligned}$$

Probabilistic model:

$$\begin{aligned} w_k &\sim \mathcal{N}(0_{n_x}, Q_k(\beta)), && k = 0, \dots, N-1, \\ v_k &\sim \mathcal{N}(0_{n_y}, R_k(\beta)), && k = 0, \dots, N, \\ x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1}) \end{aligned}$$

# Parametric model

## Parametric Linear Dynamical Model with Gaussian Noise

Dynamical model:

$$
\begin{aligned}
x_{k+1} &= A_k(\alpha)x_k + b_k(\alpha) + w_k, && k = 0, \ldots, N-1, \\
y_k &= C_k(\alpha)x_k + v_k, && k = 0, \ldots, N,
\end{aligned}
$$

Probabilistic model:

$$
\begin{aligned}
w_k &\sim \mathcal{N}(0_{n_x}, Q_k(\beta)), && k = 0, \ldots, N-1, \\
v_k &\sim \mathcal{N}(0_{n_y}, R_k(\beta)), && k = 0, \ldots, N, \\
x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})
\end{aligned}
$$

- The functions $A_k(\cdot)$, $b_k(\cdot)$ and $C_k(\cdot)$ are known functions that parameterize the model.

# Parametric model

## Parametric Linear Dynamical Model with Gaussian Noise

Dynamical model:

$$
\begin{aligned}
x_{k+1} &= A_k(\alpha)x_k + b_k(\alpha) + w_k, && k = 0, \dots, N-1, \\
y_k &= C_k(\alpha)x_k + v_k, && k = 0, \dots, N,
\end{aligned}
$$

Probabilistic model:

$$
\begin{aligned}
w_k &\sim \mathcal{N}(0_{n_x}, Q_k(\beta)), && k = 0, \dots, N-1, \\
v_k &\sim \mathcal{N}(0_{n_y}, R_k(\beta)), && k = 0, \dots, N, \\
x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})
\end{aligned}
$$

- The functions $A_k(\cdot)$, $b_k(\cdot)$ and $C_k(\cdot)$ are known functions that parameterize the model.
- The functions $Q_k(\cdot)$ and $R_k(\cdot)$ parameterize the uncertainty model.

# Parametric model

## Parametric Linear Dynamical Model with Gaussian Noise

Dynamical model:

$$x_{k+1} = A_k(\alpha)x_k + b_k(\alpha) + w_k, \qquad k = 0, \dots, N-1,$$
$$y_k = C_k(\alpha)x_k + v_k, \qquad k = 0, \dots, N,$$

Probabilistic model:

$$w_k \sim \mathcal{N}(0_{n_x}, Q_k(\beta)), \qquad k = 0, \dots, N-1,$$
$$v_k \sim \mathcal{N}(0_{n_y}, R_k(\beta)), \qquad k = 0, \dots, N,$$
$$x_0 \sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})$$

- The functions $A_k(\cdot)$, $b_k(\cdot)$ and $C_k(\cdot)$ are known functions that parameterize the model.
- The functions $Q_k(\cdot)$ and $R_k(\cdot)$ parameterize the uncertainty model.
- Prior knowledge in the form $h(\alpha, \beta) \geq 0$.

# Parametric model

## Parametric Linear Dynamical Model with Gaussian Noise

Dynamical model:

$$\begin{aligned} x_{k+1} &= A_k(\alpha)x_k + b_k(\alpha) + w_k, & k &= 0, \dots, N-1, \\ y_k &= C_k(\alpha)x_k + v_k, & k &= 0, \dots, N, \end{aligned}$$

Probabilistic model:

$$\begin{aligned} w_k &\sim \mathcal{N}(0_{n_x}, Q_k(\beta)), & k &= 0, \dots, N-1, \\ v_k &\sim \mathcal{N}(0_{n_y}, R_k(\beta)), & k &= 0, \dots, N, \\ x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1}) \end{aligned}$$

- The functions $A_k(\cdot)$, $b_k(\cdot)$ and $C_k(\cdot)$ are known functions that parameterize the model.
- The functions $Q_k(\cdot)$ and $R_k(\cdot)$ parameterize the uncertainty model.
- Prior knowledge in the form $h(\alpha, \beta) \geq 0$.

$\Rightarrow$ **Goal:** estimate the parameters $\theta := (\alpha, \beta) \in \mathbb{R}^{n_\alpha + n_\beta}$ from measurement data $y_0, \dots, y_N$.

# Typical use-case

## Parametric model for off-set free MPC

$$
\begin{aligned}
x_{k+1} &= A(u_k;\alpha)x_k + b(u_k;\alpha) + w_k^x, && k = 0,\dots,N-1,\\
d_{k+1} &= d_k + w_k^d, && k = 0,\dots,N-1,\\
y_k &= C(\alpha)x_k + d_k + v_k, && k = 0,\dots,N,
\end{aligned}
$$

$$
\begin{aligned}
\begin{bmatrix} w_k^x \\ w_k^d \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0_{n_x} \\ 0_{n_y} \end{bmatrix}, \begin{bmatrix} \beta_1 I_{n_x} & 0 \\ 0 & \beta_2 I_{n_y} \end{bmatrix}\right), && k = 0,\dots,N-1,\\
v_k &\sim \mathcal{N}\left(0_{n_y}, \beta_3 I_{n_y}\right), && k = 0,\dots,N,\\
x_0 &\sim \mathcal{N}\left(\hat{x}_{0|-1}, P_{0|-1}\right)
\end{aligned}
$$

# Typical use-case

## Parametric model for off-set free MPC

$$x_{k+1} = A(u_k;\alpha)x_k + b(u_k;\alpha) + w_k^x, \qquad k = 0,\dots,N-1,$$
$$d_{k+1} = d_k + w_k^d, \qquad k = 0,\dots,N-1,$$
$$y_k = C(\alpha)x_k + d_k + v_k, \qquad k = 0,\dots,N,$$

$$\begin{bmatrix} w_k^x \\ w_k^d \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{n_x} \\ 0_{n_y} \end{bmatrix}, \begin{bmatrix} \beta_1 I_{n_x} & 0 \\ 0 & \beta_2 I_{n_y} \end{bmatrix}\right), \qquad k = 0,\dots,N-1,$$
$$v_k \sim \mathcal{N}(0_{n_y}, \beta_3 I_{n_y}), \qquad k = 0,\dots,N,$$
$$x_0 \sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})$$

- To perform MPC, some parameter of the model often needs to be tuned.

# Typical use-case

## Parametric model for off-set free MPC

$$x_{k+1} = A(u_k;\alpha)x_k + b(u_k;\alpha) + w_k^x, \qquad k = 0,\dots,N-1,$$
$$d_{k+1} = d_k + w_k^d, \qquad k = 0,\dots,N-1,$$
$$y_k = C(\alpha)x_k + d_k + v_k, \qquad k = 0,\dots,N,$$

$$\begin{bmatrix} w_k^x \\ w_k^d \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0_{n_x} \\ 0_{n_y} \end{bmatrix}, \begin{bmatrix} \beta_1 I_{n_x} & 0 \\ 0 & \beta_2 I_{n_y} \end{bmatrix} \right), \qquad k = 0,\dots,N-1,$$
$$v_k \sim \mathcal{N}(0_{n_y}, \beta_3 I_{n_y}), \qquad k = 0,\dots,N,$$
$$x_0 \sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})$$

- To perform MPC, some parameter of the model often needs to be tuned.
- A disturbance model is also often needed to design a state estimator, espacially for offset-free MPC.

# Two examples: the random walk model

## Random walk model

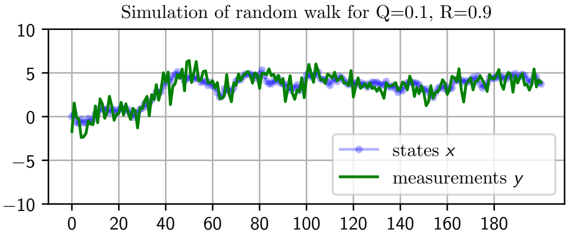
$$
\begin{aligned}
x_{k+1} &= x_k + w_k, \quad k = 0, \dots, N-1, \\
y_k &= x_k + v_k, \quad k = 0, \dots, N, \\
\\
w_k &\sim \mathcal{N}(0, q), \quad k = 0, \dots, N-1, \\
v_k &\sim \mathcal{N}(0, r), \quad k = 0, \dots, N, \\
x_0 &= 0, \\
\beta &= \begin{bmatrix} q & r \end{bmatrix}
\end{aligned}
$$

# Two examples: the random walk model

## Random walk model

$$x_{k+1} = x_k + w_k, \quad k = 0, \dots, N-1,$$
$$y_k = x_k + v_k, \quad k = 0, \dots, N,$$

$$w_k \sim \mathcal{N}(0, q), \quad k = 0, \dots, N-1,$$
$$v_k \sim \mathcal{N}(0, r), \quad k = 0, \dots, N,$$
$$x_0 = 0,$$
$$\beta = \begin{bmatrix} q & r \end{bmatrix}$$

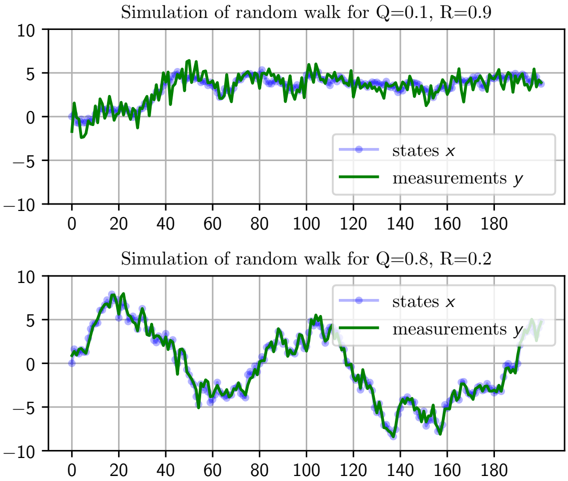


# Two examples: the random walk model

**Random walk model**

$$x_{k+1} = x_k + w_k, \quad k = 0, \dots, N-1,$$
$$y_k = x_k + v_k, \quad k = 0, \dots, N,$$

$$w_k \sim \mathcal{N}(0, q), \quad k = 0, \dots, N-1,$$
$$v_k \sim \mathcal{N}(0, r), \quad k = 0, \dots, N,$$
$$x_0 = 0,$$
$$\beta = \begin{bmatrix} q & r \end{bmatrix}$$



# Two examples: a heat transfer system

## A heat transfer system

$$
\begin{aligned}
{x_1}^+ &= (1 - {}^1/_{\tau_1})x_1 + {}^b/_{\tau_1} u + w, \\
{x_2}^+ &= (1 - {}^2/_{\tau_2})x_2 + {}^2/_{\tau_2} x_1, \\
{x_3}^+ &= (1 - {}^2/_{\tau_2})x_3 + {}^2/_{\tau_2} x_2, \\
y &= x_3 + v, \\
w &\sim \mathcal{N}(0, 10^{-3}), \\
v &\sim \mathcal{N}(0, 10^{-3}), \\
x_0 &= 0, \\
\alpha &= \begin{bmatrix} {}^1/_{\tau_1} & {}^2/_{\tau_2} & {}^b/_{\tau_1} \end{bmatrix}
\end{aligned}
$$

# Two examples: a heat transfer system

## A heat transfer system

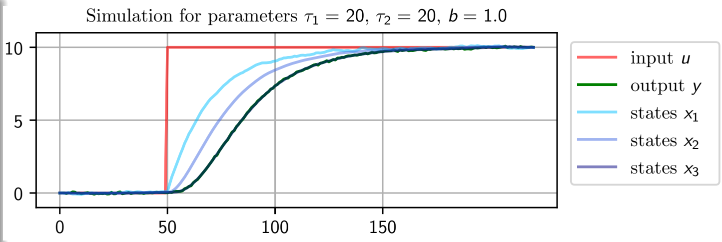
$$x_1^+ = (1 - 1/\tau_1)x_1 + b/\tau_1 u + w,$$
$$x_2^+ = (1 - 2/\tau_2)x_2 + 2/\tau_2 x_1,$$
$$x_3^+ = (1 - 2/\tau_2)x_3 + 2/\tau_2 x_2,$$
$$y = x_3 + v,$$
$$w \sim \mathcal{N}(0, 10^{-3}),$$
$$v \sim \mathcal{N}(0, 10^{-3}),$$
$$x_0 = 0,$$
$$\alpha = \begin{bmatrix} 1/\tau_1 & 2/\tau_2 & b/\tau_1 \end{bmatrix}$$

Remark: the transfer function is $G(s) = \frac{b}{1+\tau_1 s} \frac{1}{(1+\frac{\tau_2}{m} s)^m}$ with $m = 2$.

# Two examples: a heat transfer system

## A heat transfer system

$$
\begin{aligned}
{x_1}^+ &= (1 - {}^1/_{\tau_1})x_1 + {}^b/_{\tau_1} u + w, \\
{x_2}^+ &= (1 - {}^2/_{\tau_2})x_2 + {}^2/_{\tau_2} x_1, \\
{x_3}^+ &= (1 - {}^2/_{\tau_2})x_3 + {}^2/_{\tau_2} x_2, \\
y &= x_3 + v, \\
w &\sim \mathcal{N}(0, 10^{-3}), \\
v &\sim \mathcal{N}(0, 10^{-3}), \\
x_0 &= 0, \\
\alpha &= \begin{bmatrix} {}^1/_{\tau_1} & {}^2/_{\tau_2} & {}^b/_{\tau_1} \end{bmatrix}
\end{aligned}
$$

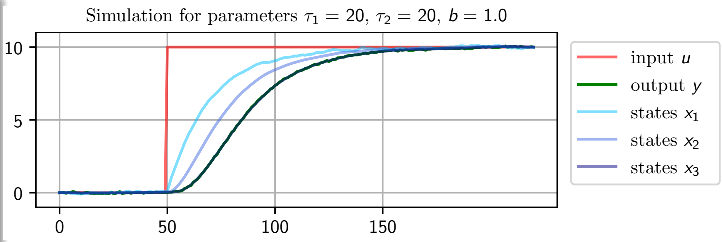


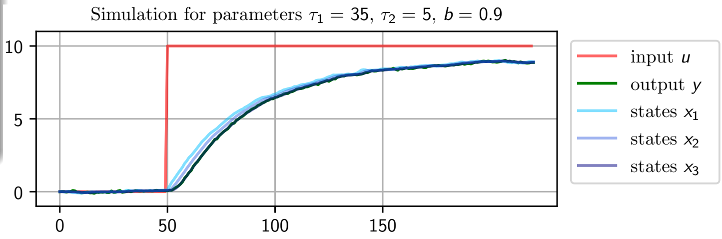

Remark: the transfer function is $G(s) = \frac{b}{1+\tau_1 s} \frac{1}{(1+\frac{\tau_2}{m}s)^m}$ with $m = 2$.

## A heat transfer system

$$x_1^+ = (1 - {}^1/_{\tau_1})x_1 + {}^b/_{\tau_1} u + w,$$
$$x_2^+ = (1 - {}^2/_{\tau_2})x_2 + {}^2/_{\tau_2} x_1,$$
$$x_3^+ = (1 - {}^2/_{\tau_2})x_3 + {}^2/_{\tau_2} x_2,$$
$$y = x_3 + v,$$
$$w \sim \mathcal{N}(0, 10^{-3}),$$
$$v \sim \mathcal{N}(0, 10^{-3}),$$
$$x_0 = 0,$$
$$\alpha = \begin{bmatrix} {}^1/_{\tau_1} & {}^2/_{\tau_2} & {}^b/_{\tau_1} \end{bmatrix}$$

Remark: the transfer function is $G(s) = \frac{b}{1+\tau_1 s} \frac{1}{(1+\frac{\tau_2}{m}s)^m}$ with $m = 2$.



# Description of the method

# The Kalman Filter

# The Kalman Filter (KF)

- A KF provides state predictions $\hat{x}_{k+1|k}, P_{k+1|k}$ given past measurements $y_0, \ldots, y_k$.
- The conditional probability law $\left(x_{k+1} \mid y_0, \ldots, y_k\right) \sim \mathcal{N}\left(\hat{x}_{k+1|k}, P_{k+1|k}\right)$ holds.

# The Kalman Filter (KF)

- A KF provides state predictions $\hat{x}_{k+1|k}, P_{k+1|k}$ given past measurements $y_0, \dots, y_k$.
- The conditional probability law $\left(x_{k+1} \mid y_0, \dots, y_k\right) \sim \mathcal{N}\left(\hat{x}_{k+1|k}, P_{k+1|k}\right)$ holds.

## The equations of the Kalman Filter

$$
\begin{aligned}
S_k &= C P_{k|k-1} C^\top + R_k(\beta), && k = 0, \dots, N, \\
K_k &= P_{k|k-1} C^\top {S_k}^{-1}, && k = 0, \dots, N, \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}), && k = 0, \dots, N, \\
P_{k|k} &= P_{k|k-1} - K_k S_k {K_k}^\top, && k = 0, \dots, N, \\
\hat{x}_{k+1|k} &= A_k(\alpha)\hat{x}_{k|k} + b_k(\alpha), && k = 0, \dots, N-1, \\
P_{k+1|k} &= A_k(\alpha) P_{k|k-1} A_k(\alpha)^\top + Q_k(\beta), && k = 0, \dots, N-1.
\end{aligned}
$$

# The Kalman Filter (KF)

- A KF provides state predictions $\hat{x}_{k+1|k}, P_{k+1|k}$ given past measurements $y_0, \ldots, y_k$.
- The conditional probability law $\left(x_{k+1} \mid y_0, \ldots, y_k\right) \sim \mathcal{N}\left(\hat{x}_{k+1|k}, P_{k+1|k}\right)$ holds.

## The equations of the Kalman Filter (different formuation)

$$
\begin{aligned}
S_k &= C P_{k|k-1} C^\top + R_k(\beta), && k = 0, \ldots, N, \\
L_k &= A_k(\alpha) P_{k|k-1} C^\top S_k{}^{-1}, && k = 0, \ldots, N, \\
\hat{x}_{k+1|k} &= \left(A_k(\alpha) - L_k C\right) \hat{x}_{k|k-1} + L_k y_k + b_k(\alpha), && k = 0, \ldots, N-1, \\
P_{k+1|k} &= A_k(\alpha) P_{k|k-1} A_k(\alpha)^\top - L_k S_k L_k{}^\top + Q_k(\beta), && k = 0, \ldots, N-1.
\end{aligned}
$$

# The Kalman Filter (KF)

- A KF provides state predictions $\hat{x}_{k+1|k}, P_{k+1|k}$ given past measurements $y_0, \ldots, y_k$.
- The conditional probability law $\left(x_{k+1} \mid y_0, \ldots, y_k\right) \sim \mathcal{N}\left(\hat{x}_{k+1|k}, P_{k+1|k}\right)$ holds.

**The equations of the Kalman Filter (different formuation)**

$$
\begin{aligned}
S_k &= C P_{k|k-1} C^\top + R_k(\beta), && k = 0, \ldots, N, \\
L_k &= A_k(\alpha) P_{k|k-1} C^\top {S_k}^{-1}, && k = 0, \ldots, N, \\
\hat{x}_{k+1|k} &= \left(A_k(\alpha) - L_k C\right) \hat{x}_{k|k-1} + L_k y_k + b_k(\alpha), && k = 0, \ldots, N-1, \\
P_{k+1|k} &= A_k(\alpha) P_{k|k-1} A_k(\alpha)^\top - L_k S_k {L_k}^\top + Q_k(\beta), && k = 0, \ldots, N-1.
\end{aligned}
$$

We define the functions "$\hat{y}_{k|k-1}(\theta) := C\hat{x}_{k|k-1}$" and "$S_k(\theta) := S_k$", with $\theta := (\alpha, \beta)$.

# The Kalman Filter (KF)

- A KF provides state predictions $\hat{x}_{k+1|k}, P_{k+1|k}$ given past measurements $y_0, \dots, y_k$.
- The conditional probability law $\left(x_{k+1} \mid y_0, \dots, y_k\right) \sim \mathcal{N}\left(\hat{x}_{k+1|k}, P_{k+1|k}\right)$ holds.

**The equations of the Kalman Filter (different formuation)**

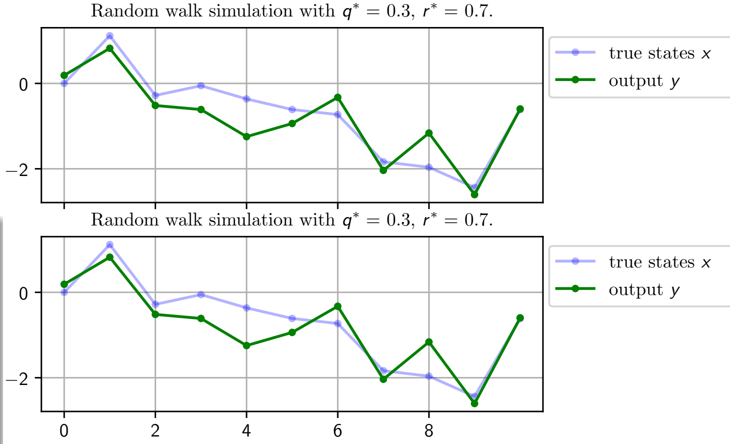
$$
\begin{aligned}
S_k &= C P_{k|k-1} C^\top + R_k(\beta), && k = 0, \dots, N, \\
L_k &= A_k(\alpha) P_{k|k-1} C^\top {S_k}^{-1}, && k = 0, \dots, N, \\
\hat{x}_{k+1|k} &= \left(A_k(\alpha) - L_k C\right) \hat{x}_{k|k-1} + L_k y_k + b_k(\alpha), && k = 0, \dots, N-1, \\
P_{k+1|k} &= A_k(\alpha) P_{k|k-1} A_k(\alpha)^\top - L_k S_k {L_k}^\top + Q_k(\beta), && k = 0, \dots, N-1.
\end{aligned}
$$

We define the functions "$\hat{y}_{k|k-1}(\theta) := C\hat{x}_{k|k-1}$" and "$S_k(\theta) := S_k$", with $\theta := (\alpha, \beta)$.

$\Rightarrow$ **Conditional probability law:** $\left(y_k \mid y_0, \dots, y_{k-1}, \theta\right) \sim \mathcal{N}\left(\hat{y}_{k|k-1}(\theta), S_k(\theta)\right)$

# The Kalman Filter in the random walk model example

- We generate data with the random walk model, with covariances $q^* = 0.3$ and $r^* = 0.7$
- We apply a KF with other values of $Q$ and $R$.

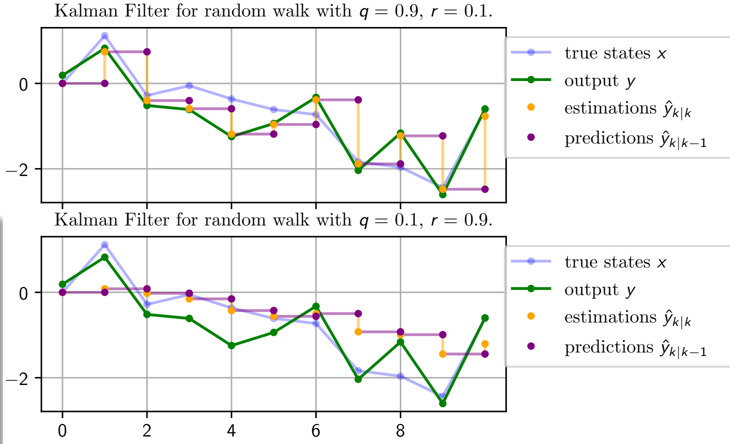



## Random walk model

$$
\begin{aligned}
x_{k+1} &= x_k + w_k, \\
y_k &= x_k + v_k, \\
w_k &\sim \mathcal{N}(0, q), \\
v_k &\sim \mathcal{N}(0, r), \\
x_0 &= 0
\end{aligned}
$$

# The Kalman Filter in the random walk model example

- We generate data with the random walk model, with covariances $q^* = 0.3$ and $r^* = 0.7$
- We apply a KF with other values of $Q$ and $R$.

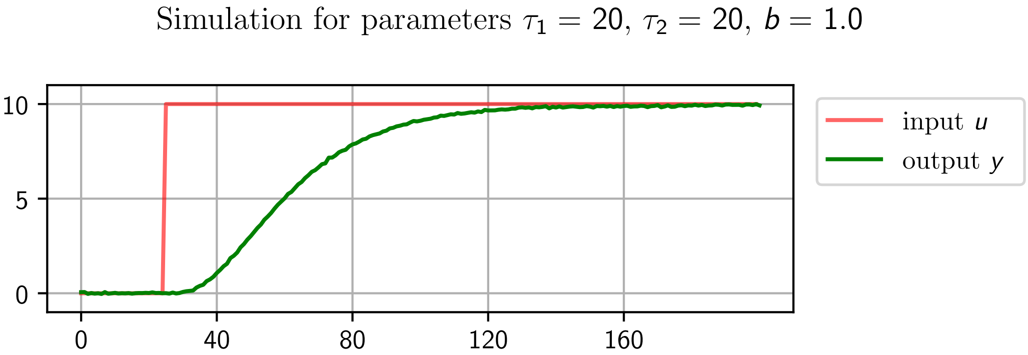



## Random walk model

$$
\begin{aligned}
x_{k+1} &= x_k + w_k, \\
y_k &= x_k + v_k, \\
w_k &\sim \mathcal{N}(0, q), \\
v_k &\sim \mathcal{N}(0, r), \\
x_0 &= 0
\end{aligned}
$$

# The Kalman Filter in the heat transfer example

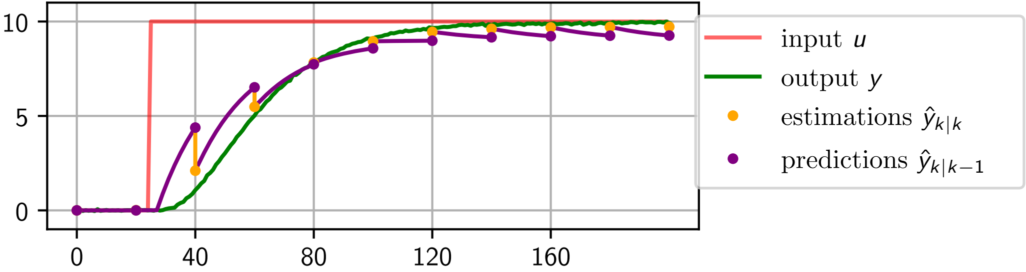
- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.



# The Kalman Filter in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

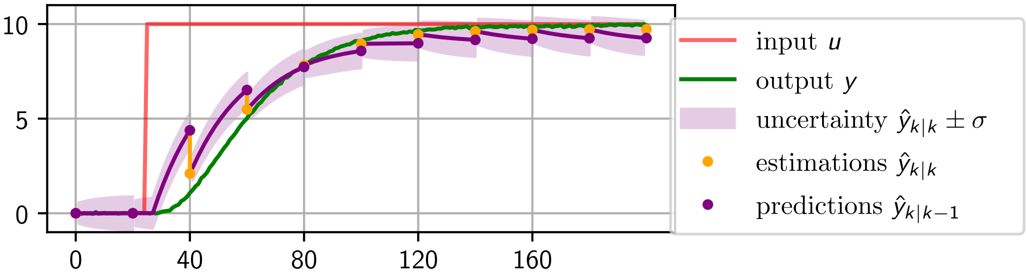
Kalman for filter for parameters $\tau_1 = 20$, $\tau_2 = 2$, $b = 0.9$



# The Kalman Filter in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

Kalman for filter for parameters $\tau_1 = 20$, $\tau_2 = 2$, $b = 0.9$



# The optimization problem

# Qualitative description

- One can apply a Kalman Filter to the measurement data for estimated parameters $\alpha, \beta$.

# Qualitative description

- One can apply a Kalman Filter to the measurement data for estimated parameters $\alpha, \beta$.
- We seek for the parameters resulting in the "best KF".

# Qualitative description

- One can apply a Kalman Filter to the measurement data for estimated parameters $\alpha, \beta$.
- We seek for the parameters resulting in the "best KF".
- We measure the quality of the KF with the prediction error $y_k - C\hat{x}_{k|k-1}$.

# Qualitative description

- One can apply a Kalman Filter to the measurement data for estimated parameters $\alpha, \beta$.
- We seek for the parameters resulting in the "best KF".
- We measure the quality of the KF with the prediction error $y_k - C\hat{x}_{k|k-1}$.
- This can be refined by considering not only the prediction error, but also its estimated covariance $S_k$.

# Qualitative description

- One can apply a Kalman Filter to the measurement data for estimated parameters $\alpha, \beta$.
- We seek for the parameters resulting in the "best KF".
- We measure the quality of the KF with the prediction error $y_k - C\hat{x}_{k|k-1}$.
- This can be refined by considering not only the prediction error, but also its estimated covariance $S_k$.
- This method belongs to the class of *prediction error estimation methods* [2], formulated for a state-space model [3].

[2] L. Ljung, "Prediction error estimation methods," *Circuits, Systems and Signal Processing, vol. 21, no. 1, pp. 11–21*, 2002.

[3] J.Valluru, P. Lakhmani, S.C. Patwardhan and L.T. Biegler, "Development of moving window state and parameter estimators under maximum likelihood and Bayesian frameworks," *Journal of Process Control, vol. 60, pp. 48-67*, 2017

# A first optimization problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \sum_{k=0}^{N} \left\| y_k - \hat{y}_{k|k-1}(\theta) \right\|^2 \\ \text{subject to} \quad & h(\theta) \geq 0. \end{aligned}$$

# A first optimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{k=0}^{N} \left\| y_k - \hat{y}_{k|k-1}(\theta) \right\|^2$$
$$\text{subject to} \quad h(\theta) \geq 0.$$

**Lifted form:**

$$\begin{aligned}
\underset{\alpha,\beta,\boldsymbol{S},\boldsymbol{L},\boldsymbol{x},\boldsymbol{P}}{\text{minimize}} \quad & \sum_{k=0}^{N} \left\| y_k - C\hat{x}_{k|k-1} \right\|^2 \\
\text{subject to} \quad & S_k = CP_{k|k-1}C^\top + R_k(\beta), && k = 0,\dots,N, \\
& L_k = A_k(\alpha)P_{k|k-1}C^\top S_k{}^{-1}, && k = 0,\dots,N, \\
& \hat{x}_{k+1|k} = (A_k(\alpha) - CL_k)\hat{x}_{k|k-1} + L_k y_k + b_k(\alpha), && k = 0,\dots,N-1, \\
& P_{k+1|k} = A_k(\alpha)P_{k|k-1}A_k(\alpha)^\top - L_k S_k L_k{}^\top + Q_k(\beta), && k = 0,\dots,N-1, \\
& h(\alpha,\beta) \geq 0.
\end{aligned}$$

# A more accurate optimization problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \sum_{k=0}^{N} \left\| y_k - \hat{y}_{k|k-1}(\theta) \right\|^2_{S_k(\theta)^{-1}} + \log \left| S_k(\theta) \right| \\ \text{subject to} \quad & h(\theta) \geq 0. \end{aligned}$$

# A more accurate optimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{k=0}^{N} \left\| y_k - \hat{y}_{k|k-1}(\theta) \right\|^2_{S_k(\theta)^{-1}} + \log |S_k(\theta)|$$

$$\text{subject to} \quad h(\theta) \geq 0.$$

<u>Notations:</u>

$$\|x\|_M := x^\top M x$$

$$|M| := \det(M)$$

# A more accurate optimization problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \sum_{k=0}^{N} \left\| y_k - \hat{y}_{k|k-1}(\theta) \right\|^2_{S_k(\theta)^{-1}} + \log |S_k(\theta)| \\ \text{subject to} \quad & h(\theta) \geq 0. \end{aligned}$$

**Notations:**

$$\|x\|_M := x^\top M x$$
$$|M| := \det(M)$$

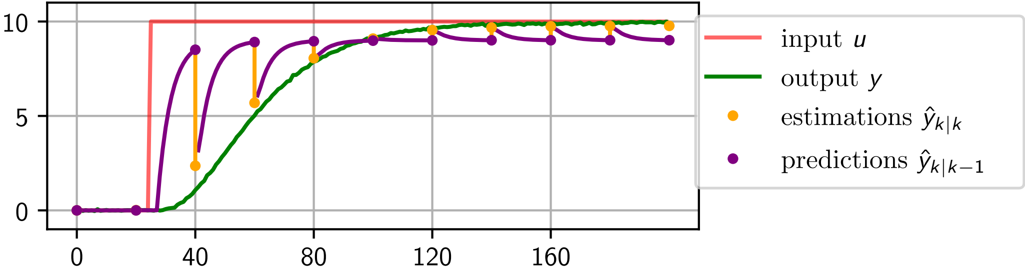
**Lifted form:**

$$\begin{aligned} \underset{\alpha,\beta,\boldsymbol{S},\boldsymbol{L},\boldsymbol{x},\boldsymbol{P}}{\text{minimize}} \quad & \sum_{k=0}^{N} \left\| y_k - C\hat{x}_{k|k-1} \right\|^2_{S_k^{-1}} + \log |S_k| \\ \text{subject to} \quad & S_k = C P_{k|k-1} C^\top + R_k(\beta), && k = 0, \ldots, N, \\ & L_k = A_k(\alpha) P_{k|k-1} C^\top S_k{}^{-1}, && k = 0, \ldots, N, \\ & \hat{x}_{k+1|k} = (A_k(\alpha) - L_k C)\hat{x}_{k|k-1} + L_k y_k + b_k(\alpha), && k = 0, \ldots, N-1, \\ & P_{k+1|k} = A_k(\alpha) P_{k|k-1} A_k(\alpha)^\top - L_k S_k L_k{}^\top + Q_k(\beta), && k = 0, \ldots, N-1, \\ & h(\alpha, \beta) \geq 0. \end{aligned}$$

# Some optimization steps in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

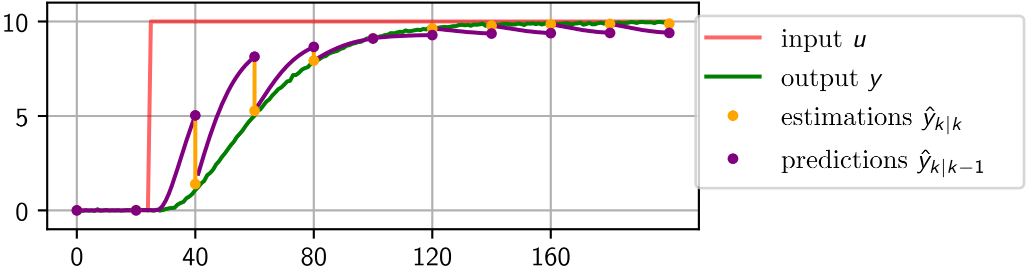
Kalman for filter for parameters $\tau_1 = 5.0$, $\tau_2 = 2.0$, $b = 0.9$



# Some optimization steps in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

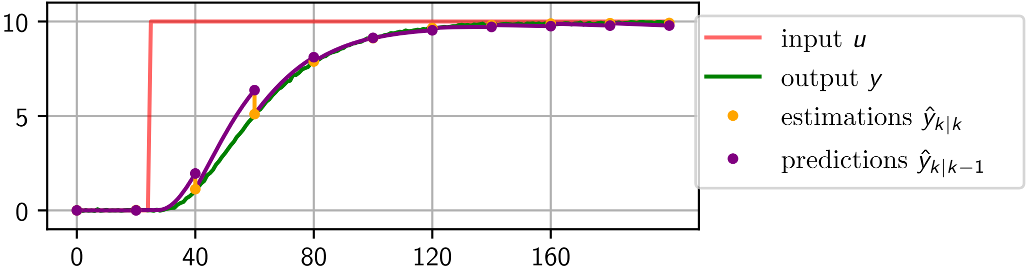
Kalman for filter for parameters $\tau_1 = 9.5$, $\tau_2 = 7.4$, $b = 0.9$



# Some optimization steps in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

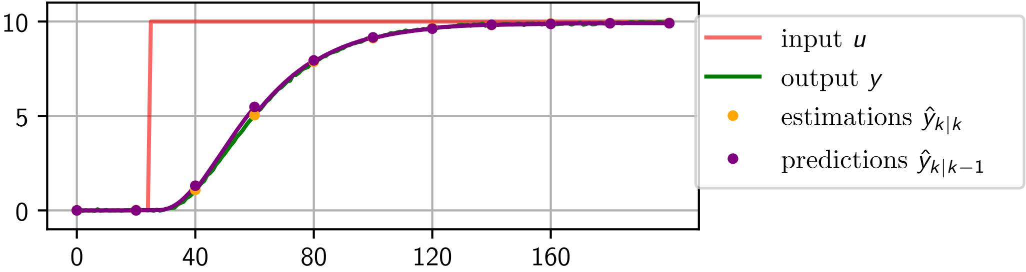
Kalman for filter for parameters $\tau_1 = 15.5$, $\tau_2 = 14.6$, $b = 1.0$



# Some optimization steps in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

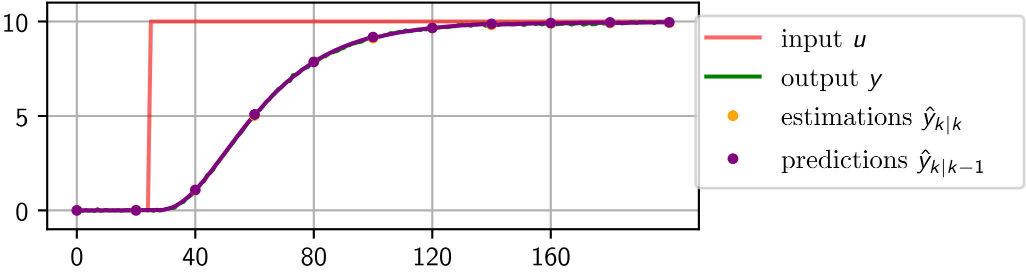
Kalman for filter for parameters $\tau_1 = 18.5$, $\tau_2 = 18.2$, $b = 1.0$



# Some optimization steps in the heat transfer example

- We generate data with the heat transfer model, with parameters $\tau_1^* = 20$ and $\tau_2^* = 20$ and $b^* = 1.0$
- We apply a KF with other values of $\tau_1$ and $\tau_2$ and $b$.

Kalman for filter for parameters $\tau_1 = 20.0$, $\tau_2 = 20.0$, $b = 1.0$



# Relation with Maximum Likelihood Estimation

# Relation with maximum likelihood estimation

## Maximum likelihood estimation

**Theorem:** The former optimization problem is equivalent to the following

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p\left(y_0, \ldots, y_N \mid \theta\right) \\ \text{subject to} \quad & h(\theta) \geq 0 \end{aligned}$$

# Relation with maximum likelihood estimation

## Maximum likelihood estimation

**Theorem:** The former optimization problem is equivalent to the following

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p\left(y_0, \ldots, y_N \mid \theta\right) \\ \text{subject to} \quad & h(\theta) \geq 0 \end{aligned}$$

Remark :

For a given prior knowledge on $\theta$, this can easily be related to the Maximum A posteriori Probability (MAP):

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p\left(\theta \mid y_0, \ldots, y_N\right) \qquad = p\left(\theta\right) \times p\left(y_0, \ldots, y_N \mid \theta\right) \\ \text{subject to} \quad & h(\theta) \geq 0 \end{aligned}$$

# Relation with maximum likelihood estimation

## Maximum likelihood estimation

**Theorem:** The former optimization problem is equivalent to the following

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p\left(y_0, \ldots, y_N \mid \theta\right) \\ \text{subject to} \quad & h(\theta) \geq 0 \end{aligned}$$

Remark :

For a given prior knowledge on $\theta$, this can easily be related to the Maximum A posteriori Probability (MAP):

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p\left(\theta \mid y_0, \ldots, y_N\right) \qquad = p\left(\theta\right) \times p\left(y_0, \ldots, y_N \mid \theta\right) \\ \text{subject to} \quad & h(\theta) \geq 0 \end{aligned}$$

Furthermore, these two problems are equivalent for a *non-informative prior* i.e. a prior with uniform distribution on the set $\{\theta \in \mathbb{R}^{n_\alpha + n_\beta} \quad \text{such that} \quad h(\theta) \geq 0\}$ (if bounded)

# Relation with maximum likelihood estimation

## Maximum likelihood estimation

**Theorem:** The former optimization problem is equivalent to the following

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p\left(y_0, \ldots, y_N \mid \theta\right) \\ \text{subject to} \quad & h(\theta) \geq 0 \end{aligned}$$

*Proof :*

$$p\left(y_0, \ldots, y_N \mid \theta\right) = \prod_{k=0}^{N} p\left(y_k \mid y_0, \ldots, y_{k-1}, \theta\right) = \prod_{k=0}^{N} f_{\text{Gauss}}\left(y_k; \hat{y}_{k|k-1}(\theta), S_k(\theta)\right),$$

with $f_{\text{Gauss}}(x; \mu, S) =: (2\pi\,|S|)^{-1/2}\, e^{-\frac{1}{2}\|x-\mu\|_{S^{-1}}}$. Hence the following holds

$$-2\log\left(p\left(y_0, \ldots, y_N \mid \theta\right)\right) = \sum_{k=0}^{N} \left\|y_k - \hat{y}_{k|k-1}(\theta)\right\|^2_{S_k(\theta)^{-1}} + \log\left|S_k(\theta)\right| + (N+1)n_y \log(2\pi)$$

# Comparison with Trajectory Optimization

# Comparison with Trajectory Optimization

### Prediction error methods

$$\underset{\alpha,\beta,\boldsymbol{S},\boldsymbol{L},\boldsymbol{x},\boldsymbol{P}}{\text{minimize}} \sum_{k=0}^{N} \left\| y_k - C\hat{x}_{k|k-1} \right\|^2_{S_k^{-1}} + \log |S_k|$$

subject to

$$S_k = CP_{k|k-1}C^\top + R(\beta),$$

$$L_k = A_k(\alpha)P_{k|k-1}C^\top S_k{}^{-1},$$

$$\hat{x}_{k+1|k} = (A_k(\alpha) - L_k C)\hat{x}_{k|k-1} + L_k y_k + b_k(\alpha),$$

$$P_{k+1|k} = A_k(\alpha)P_{k|k}A_k(\alpha)^\top - L_k S_k L_k{}^\top + Q(\beta),$$

$$h(\alpha, \beta) \geq 0.$$

### Trajectory optimization methods

$$\underset{\alpha,\boldsymbol{x},\boldsymbol{w},\boldsymbol{v}}{\text{minimize}} \quad \sum_{k=0}^{N} \| w_k \|^2_{Q(\beta)^{-1}} + \| v_k \|^2_{R(\beta)^{-1}}$$

subject to

$$x_{k+1} = A_k(\alpha)x_k + b_k(\alpha) + w_k,$$

$$y_k = Cx_k + v_k,$$

$$h(\alpha, \beta) \geq 0.$$

# Comparison with Trajectory Optimization

## Prediction error methods

**Pros**

- Can find the noise covariances $Q$, $R$,
- *Almost surely convergence* theorems,
- Is the maximum likelihood estimator,
- "Single shooting" formulation is possible

**Cons**

- Designed for linear systems.
- State or disturbance constraints are impossible.

**Derived from**

$$\underset{\theta}{\text{maximize}} \quad p\left(\theta \mid y_0, \ldots, y_N\right)$$

## Trajectory optimization methods

**Pros**

- State or disturbance constraints are possible.
- Designed for linear or nonlinear systems.
- Stability theorems

**Cons**

- Require $Q$ and $R$ as prior knowledge,
- When the probabilistic aspect is significant, sometime fail to estimate some parameters.

**Derived from**

$$\underset{\boldsymbol{x}, \theta}{\text{maximize}} \quad p\left(x_0, \ldots, x_N, \theta \mid y_0, \ldots, y_N\right)$$

# Counter example with Trajectory Optimization

**Why would Trajectory Optimization fail to estimate $\beta$ ?**

**Parametric model for off-set free MPC**

$$
\begin{aligned}
x^+ &= A(u;\alpha)x + b(u;\alpha) + w_x, \\
d^+ &= d + w_d, \\
y &= C(\alpha)x + d + v, \\
w_x &\sim \mathcal{N}\left(0_{n_x}, \sigma_x^2 I_{n_x}\right), \\
w_d &\sim \mathcal{N}\left(0_{n_y}, {\sigma_d}^2 I_{n_y}\right), \\
v &\sim \mathcal{N}\left(0_{n_y}, \sigma_y^2 I_{n_y}\right).
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
x^+ &= A(u;\alpha)x + b(u;\alpha) + w_x, \\
\tilde{d}^+ &= \tilde{d} + \tilde{w}_d, \\
y &= C(\alpha)x + \sigma_d \tilde{d} + v, \\
w_x &\sim \mathcal{N}\left(0_{n_x}, \sigma_x^2 I_{n_x}\right) \\
\tilde{w}_d &\sim \mathcal{N}\left(0_{n_y}, I_{n_y}\right) \\
v &\sim \mathcal{N}\left(0_{n_y}, \sigma_y^2 I_{n_y}\right).
\end{aligned}
$$

CdV:

$$
\tilde{d} = {\sigma_d}^{-1} d, \quad \tilde{w}_d = {\sigma_d}^{-1} w_d
$$

# Counter example with Trajectory Optimization

**Why would Trajectory Optimization fail to estimate $\beta$ ?**

**Parametric model for off-set free MPC**

$$
\begin{aligned}
x^+ &= A(u;\alpha)x + b(u;\alpha) + w_x, \\
d^+ &= d + w_d, \\
y &= C(\alpha)x + d + v, \\
w_x &\sim \mathcal{N}\left(0_{n_x}, \sigma_x^2 I_{n_x}\right), \\
w_d &\sim \mathcal{N}\left(0_{n_y}, {\sigma_d}^2 I_{n_y}\right), \\
v &\sim \mathcal{N}\left(0_{n_y}, \sigma_y^2 I_{n_y}\right).
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
x^+ &= A(u;\alpha)x + b(u;\alpha) + w_x, \\
\tilde{d}^+ &= \tilde{d} + \tilde{w}_d, \\
y &= C(\alpha)x + \sigma_d \tilde{d} + v, \\
w_x &\sim \mathcal{N}\left(0_{n_x}, \sigma_x^2 I_{n_x}\right) \\
\tilde{w}_d &\sim \mathcal{N}\left(0_{n_y}, I_{n_y}\right) \\
v &\sim \mathcal{N}\left(0_{n_y}, \sigma_y^2 I_{n_y}\right).
\end{aligned}
$$

<u>CdV:</u>

$$
\begin{aligned}
\tilde{d} &= {\sigma_d}^{-1} d, \\
\tilde{w}_d &= {\sigma_d}^{-1} w_d
\end{aligned}
$$

$\Rightarrow$ **Solution of Trajectory optimization**:

$$\sigma_d \tilde{d} \approx y \qquad \tilde{d}^+ - \tilde{d} \approx 0 \qquad \sigma_d \approx +\infty$$

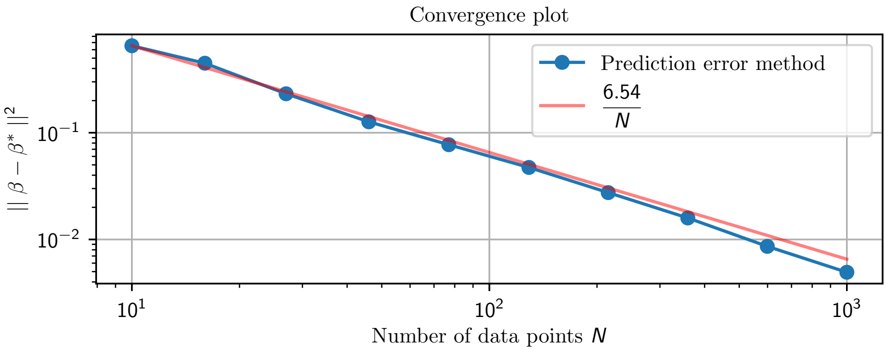
# A small benchmark

# Convergence of the estimator in the random walk model example

- We generate 100 of different values of $\beta^* := (q^*, r^*)$.
- For each sample, we simulate the random walk model to get measurements $y_0, \ldots, y_N$.
- For each sample, we use our method to estimate $\beta$ from the measurements
- We compute the MSE, i.e. $\|\beta - \beta^*\|^2$ over the sample.
- We repeat these for different values of $N$.

# Convergence of the estimator in the random walk model example

- We generate 100 of different values of $\beta^* := (q^*, r^*)$.
- For each sample, we simulate the random walk model to get measurements $y_0, \ldots, y_N$.
- For each sample, we use our method to estimate $\beta$ from the measurements
- We compute the MSE, i.e. $\|\beta - \beta^*\|^2$ over the sample.
- We repeat these for different values of $N$.

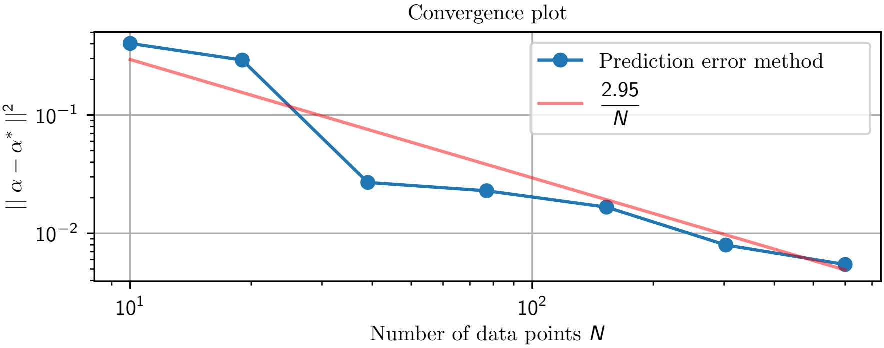


# Convergence of the estimator in the heat transfer example

- We repeat the same procedure for the heat transfer system, for $\alpha := \begin{bmatrix} 1/\tau_1 & 2/\tau_2 & b/\tau_1 \end{bmatrix}$
- We only generate 40 values of $\alpha$.

# Convergence of the estimator in the heat transfer example

- We repeat the same procedure for the heat transfer system, for $\alpha := \begin{bmatrix} 1/\tau_1 & 2/\tau_2 & b/\tau_1 \end{bmatrix}$
- We only generate 40 values of $\alpha$.



# Open questions

# An open question of optimization

**How to keep maximum degrees of freedom regarding $Q(\cdot)$ and $R(\cdot)$ ?**

# An open question of optimization

**How to keep maximum degrees of freedom regarding $Q(\cdot)$ and $R(\cdot)$ ?**

More precisely, how to deal with the following type of optimization problem ?

$$\begin{aligned} \underset{\alpha, L, S, P}{\text{minimize}} \quad & F(\alpha, L, S) \\ \text{subject to} \quad & P \succ 0, \\ & L = A(\alpha) P C^\top S^{-1}, \\ & S \succ C P C^\top, \\ & P + L S L^\top \succ A(\alpha) P A(\alpha)^\top \end{aligned}$$

# An open question of MPC Stability

## Parametric model for off-set free MPC

$$\begin{aligned} x^+ &= A(u;\alpha)x + b(u;\alpha) + w_x, \\ d^+ &= d + w_d, \\ y &= C(\alpha)x + d + v, \\ w &\sim \mathcal{N}\left(0_{n_x}, Q(\beta)\right), \\ v &\sim \mathcal{N}\left(0_{n_y}, R(\beta)\right). \end{aligned}$$

# An open question of MPC Stability

## Parametric model for off-set free MPC

$$\begin{aligned} x^+ &= A(u;\alpha)x + b(u;\alpha) + w_x, \\ d^+ &= d + w_d, \\ y &= C(\alpha)x + d + v, \\ w &\sim \mathcal{N}\left(0_{n_x}, Q(\beta)\right), \\ v &\sim \mathcal{N}\left(0_{n_y}, R(\beta)\right). \end{aligned}$$

- The presented model provides an estimate for the parameter $\beta$ that will provide the **best predictions**.
- What we actually want is an estimate for $\beta$ that will provide the **best closed-loop control performance.**
- For example, $\mathrm{Cov}(w_k^d)$ plays an important role: if it is too high, the controller might destabilize the system, but if it is 0, offsets might remain in regulation problems.

# The important particular case of Linear Time Invariant (LTI) systems

# The important particular case of LTI systems

## Parametric LTI with Gaussian Noise

$$
\begin{aligned}
x_{k+1} &= A(\alpha)x_k + b_k(\alpha) + w_k, && k = 0, \dots, N-1,\\
y_k &= Cx_k + v_k, && k = 0, \dots, N,\\
w_k &\sim \mathcal{N}(0_{n_x}, Q(\beta)), && k = 0, \dots, N-1,\\
v_k &\sim \mathcal{N}(0_{n_y}, R(\beta)), && k = 0, \dots, N,\\
x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})
\end{aligned}
$$

# The important particular case of LTI systems

## Parametric LTI with Gaussian Noise

$$
\begin{aligned}
x_{k+1} &= A(\alpha)x_k + b_k(\alpha) + w_k, && k = 0, \dots, N-1, \\
y_k &= Cx_k + v_k, && k = 0, \dots, N, \\
w_k &\sim \mathcal{N}(0_{n_x}, Q(\beta)), && k = 0, \dots, N-1, \\
v_k &\sim \mathcal{N}(0_{n_y}, R(\beta)), && k = 0, \dots, N, \\
x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1})
\end{aligned}
$$

In this case, the KF will quickly converge to its stationary state.

# The important particular case of LTI systems

## Parametric LTI with Gaussian Noise

$$\begin{aligned} x_{k+1} &= A(\alpha)x_k + b_k(\alpha) + w_k, & k &= 0,\dots,N-1, \\ y_k &= Cx_k + v_k, & k &= 0,\dots,N, \\ w_k &\sim \mathcal{N}(0_{n_x}, Q(\beta)), & k &= 0,\dots,N-1, \\ v_k &\sim \mathcal{N}(0_{n_y}, R(\beta)), & k &= 0,\dots,N, \\ x_0 &\sim \mathcal{N}(\hat{x}_{0|-1}, P_{0|-1}) & & \end{aligned}$$

In this case, the KF will quickly converge to its stationary state.

## The discrete algebraic Ricatti equation

$$\begin{aligned} S &= CPC^\top + R(\beta), \\ L &= A(\alpha)PC^\top S^{-1}, \\ P &= A(\alpha)PA(\alpha)^\top - LSL^\top + Q(\beta). \end{aligned}$$

# The important particular case of LTI systems

By approximating the Kalman Filter equations with their stationary equivalents, the method boils down to:

## Prediction error method for LTI systems

$$
\begin{aligned}
\underset{\alpha,\beta,S,L,P,\boldsymbol{x}}{\text{minimize}} \quad & \frac{1}{N+1}\sum_{k=0}^{N} \left\| y_k - C\hat{x}_{k|k-1} \right\|_{S^{-1}}^2 + \log|S| \\
\text{subject to} \quad & \hat{x}_{k+1|k} = (A(\alpha) - LC)\hat{x}_{k|k-1} + Ly_k + b_k(\alpha), \qquad k = 0, \ldots, N-1, \\
& S = CPC^\top + R(\beta), \\
& L = A(\alpha)PC^\top S^{-1}, \\
& P = A(\alpha)PA(\alpha)^\top - LSL^\top + Q(\beta), \\
& P \succ 0, \\
& h(\alpha, \beta) \geq 0.
\end{aligned}
$$

# The particular case of single-output LTI

When $n_y = 1$, more simplifications can be made. Especially we let maximum degrees of freedom in $Q(\cdot)$ and $R(\cdot)$:

## Prediction error method for single-output LTI systems

$$\begin{aligned}
\underset{\alpha, L, P, \boldsymbol{x}}{\text{minimize}} \quad & \sum_{k=0}^{N} (y_k - C\hat{x}_{k|k-1})^2 \\
\text{subject to} \quad & \hat{x}_{k+1|k} = (A(\alpha) - LC)\hat{x}_{k|k-1} + Ly_k + b_k(\alpha), \qquad k = 0, \ldots, N-1, \\
& h(\alpha) \geq 0, \\
& L = A(\alpha)PC^\top, \\
& P \succ 0, \\
& P - A(\alpha)PA(\alpha)^\top + LSL^\top \succ 0, \\
& CPC^\top \leq 1.
\end{aligned}$$

# An open question of optimization

**How to keep maximum degrees of freedom regarding $Q(\cdot)$ and $R(\cdot)$ ?**

# An open question of optimization

**How to keep maximum degrees of freedom regarding $Q(\cdot)$ and $R(\cdot)$ ?**

More precisely, how to deal with the following type of optimization problem ?

$$
\begin{aligned}
\underset{\alpha, L, P}{\text{minimize}} \quad & F(\alpha, L) \\
\text{subject to} \quad & P \succ 0, \\
& L = A(\alpha) P C^\top, \\
& C P C^\top \leq 1, \\
& P + L L^\top \succ A(\alpha) P A(\alpha)^\top
\end{aligned}
$$