How to estimate parameters of linear system using noisy measurement data?

Léo Simpson

Overview

An industrial problem

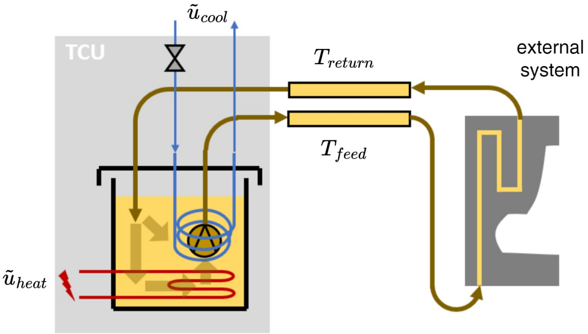
Temperature Control Units

Temperature Control Units (TCU) are machines that aim to control the temperature of a circulating fluid for the purpose of controlling the temperature of another device.

Temperature Control Units

Temperature Control Units (TCU) are machines that aim to control the temperature of a circulating fluid for the purpose of controlling the temperature of another device.

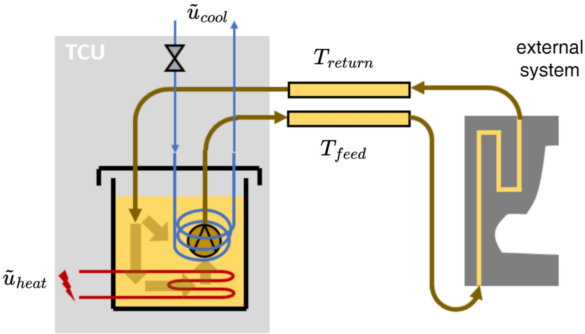
- It uses cooling water and electrical resistances
- Context: a new control system is being implemented.



Relation between the internal and external systems.

- Goal: control the temperature of the external system, using the actuators of the TCU.

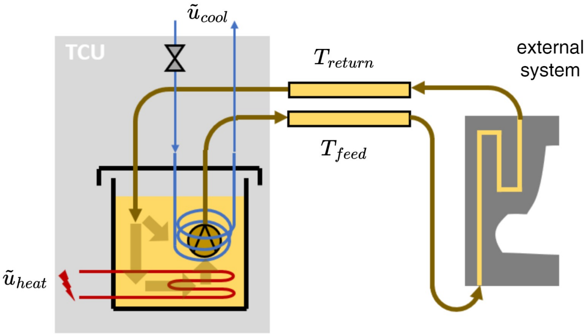
System to control : TCU + exterior



Relation between the internal and external systems.

- Goal: control the temperature of the external system, using the actuators of the TCU.
- The internal system is modelled using **physics and experimental data**.

System to control : TCU + exterior

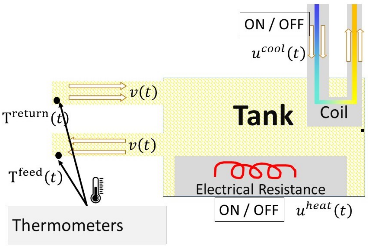


Relation between the internal and external systems.

- Goal: control the temperature of the external system, using the actuators of the TCU.
- The internal system is modelled using **physics and experimental data**.
- The exterior system is unknown and changing, hence it has to be learned online.

A modelling question

Using physics knowledge about the system, we make a very simple model



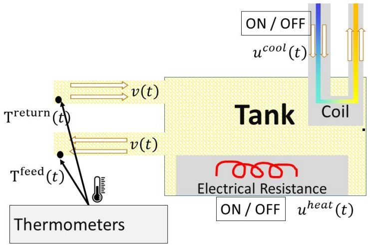
Simplified physics of a TCU

$$
\begin{aligned}
\dot{T}^{\text{tank}} &= h^{\text{cool}} u^{\text{cool}} (T^{\text{cool}} - T^{\text{tank}}) + \\
&\quad Q^{\text{heat}} u^{\text{heat}} + \\
&\quad h^{\text{ambient}} (T^{\text{ambient}} - T^{\text{tank}}) + \\
&\quad v h^{\text{return}} (T^{\text{return}} - T^{\text{tank}}), \\
\dot{T}^{\text{feed}} &= v h^{\text{feed}} (T^{\text{tank}} - T^{\text{feed}}), \\
\dot{T}^{\text{ambient}} &= 0, \\
T^{\text{feed}} &= T^{\text{feed}}
\end{aligned}
$$

- The state of the system is $x := \begin{pmatrix} T^{\text{tank}} & T^{\text{feed}} & T^{\text{ambient}} \end{pmatrix}$
- The input is $u := \begin{pmatrix} v & u^{\text{cool}} & u^{\text{cool}} & T^{\text{return}} \end{pmatrix}$ and the output is $y := T^{\text{feed}}(t)$
- The unknown parameters are $\theta := \begin{pmatrix} h^{\text{cool}} & Q^{\text{heat}} & h^{\text{ambient}} & h^{\text{return}} & h^{\text{feed}} \end{pmatrix}$

A modelling question

If the sampling time of the data is constant, it is more convenient to learn a discrete-time system:



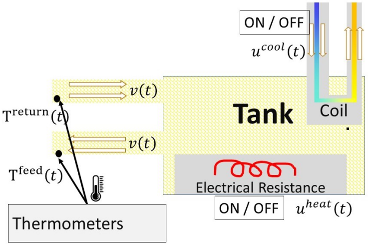
Simplified physics of a TCU

$$
\begin{aligned}
T^{\text{tank}}(t+1) &= T^{\text{tank}}(t) + Q^{\text{heat}} u^{\text{heat}}(t) + \\
&\quad + h^{\text{cool}}(t) u^{\text{cool}}(T^{\text{cool}} - T^{\text{tank}}(t)) \\
&\quad h^{\text{ambient}}(T^{\text{ambient}}(t) - T^{\text{tank}}(t)) + \\
&\quad v(t) h^{\text{return}}(T^{\text{return}}(t) - T^{\text{tank}}(t)), \\
T^{\text{feed}}(t+1) &= T^{\text{feed}}(t) + v(t) h^{\text{feed}}(T^{\text{tank}}(t) - T^{\text{feed}}(t)), \\
T^{\text{amb.}}(t+1) &= T^{\text{amb.}}(t), \\
T^{\text{feed}}(t) &= T^{\text{feed}}(t)
\end{aligned}
$$

- The state of the system is $x := \begin{pmatrix} T^{\text{tank}} & T^{\text{feed}} & T^{\text{ambient}} \end{pmatrix}$
- The input is $u := \begin{pmatrix} v & u^{\text{cool}} & u^{\text{cool}} & T^{\text{return}} \end{pmatrix}$ and the output is $y := T^{\text{feed}}(t)$
- The unknown parameters are $\theta := \begin{pmatrix} h^{\text{cool}} & Q^{\text{heat}} & h^{\text{ambient}} & h^{\text{return}} & h^{\text{feed}} \end{pmatrix}$

A modelling question

A model is never perfect, and disturbance are usually present

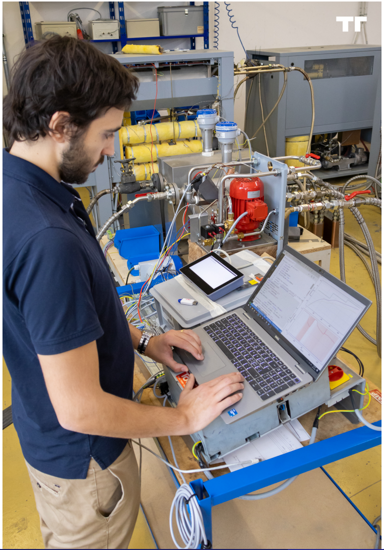


Simplified physics of a TCU

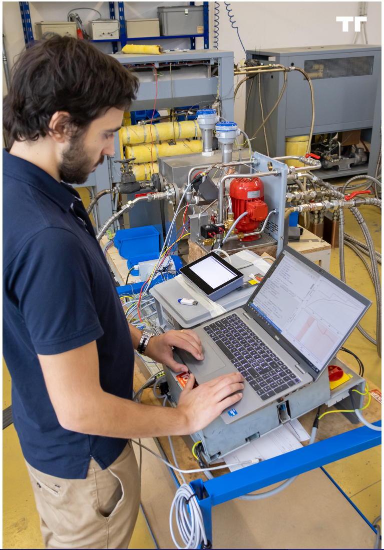
$$
\begin{aligned}
T^{\text{tank}}(t+1) &= T^{\text{tank}}(t) + h^{\text{cool}} u^{\text{cool}}(t)(T^{\text{cool}} - T^{\text{tank}}(t)) + \\
&\quad Q^{\text{heat}} u^{\text{heat}}(t) + \\
&\quad h^{\text{ambient}}(T^{\text{ambient}}(t) - T^{\text{tank}}(t)) + \\
&\quad v(t) h^{\text{return}}(T^{\text{return}}(t) - T^{\text{tank}}(t)) + w^{\text{tank}}(t), \\
T^{\text{feed}}(t+1) &= T^{\text{feed}}(t) + v(t) h^{\text{feed}}(T^{\text{tank}}(t) - T^{\text{feed}}(t)), \\
T^{\text{ambient}}(t+1) &= T^{\text{ambient}}(t) + w^{\text{ambient}}(t), \\
T^{\text{feed}}(t) &= T^{\text{feed}}(t) + v(t)
\end{aligned}
$$

- The state of the system is $x := \begin{pmatrix} T^{\text{tank}} & T^{\text{feed}} & T^{\text{ambient}} \end{pmatrix}$
- The input is $u := \begin{pmatrix} v & u^{\text{cool}} & u^{\text{cool}} & T^{\text{return}} \end{pmatrix}$ and the output is $y := T^{\text{feed}}(t)$
- The unknown parameters are $\theta := \begin{pmatrix} h^{\text{cool}} & Q^{\text{heat}} & h^{\text{ambient}} & h^{\text{return}} & h^{\text{feed}} \end{pmatrix}$

A System Identification question

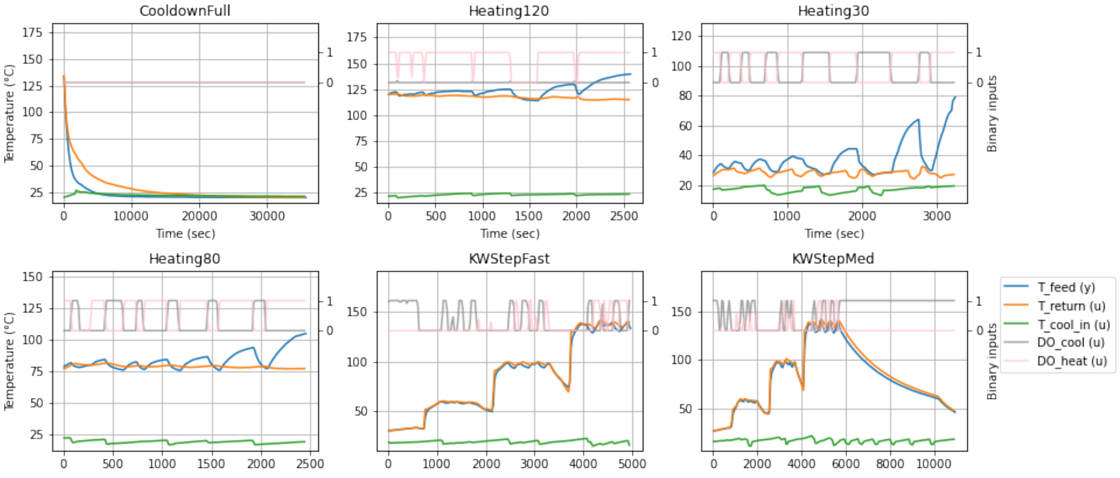


A System Identification question



- Datasets have been collected by conducting experiments on the machines
- Each dataset contains a sequence of input together with a sequence of outputs
- How to use this data to infer the parameters of our physics-based model?

Experimental data

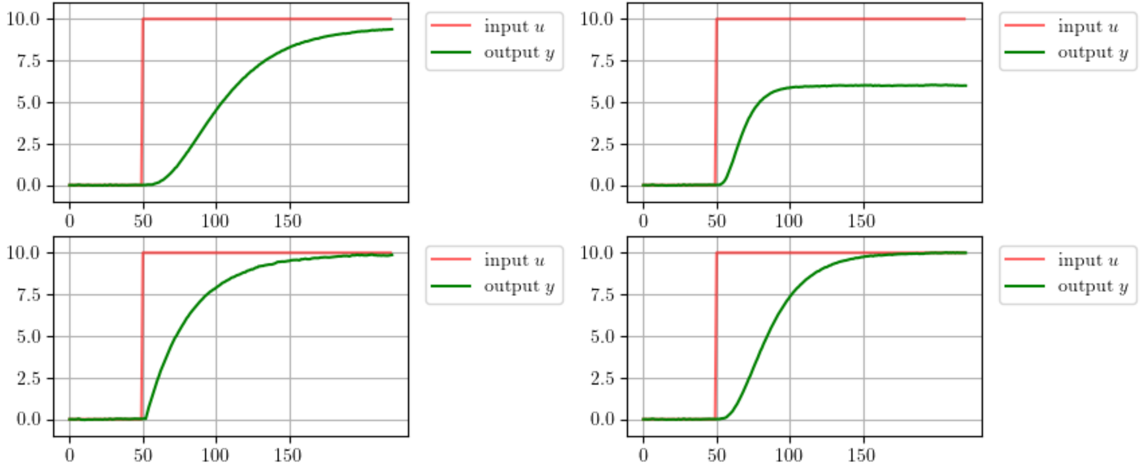


6 datasets collected from experiments on one of the TCU.
(in total, 16 datasets are available for more than 7 machines)

System Identification Task

- This shows the need for an efficient method to estimate parameters of a linear dynamical system.
- Estimating the deterministic part together with the noise model also has the advantage to design a good online state estimator.
- Such an algorithm could also be used online, to learn the external system.

Typical external systems



Impulse responses for some examples of external heat transfer systems

Problem Statement

Parametric Linear Time-Variant models

Parametric Dynamical Models with Random Noise

$$
\begin{aligned}
x_{k+1} &= f\left(x_k,\ u_k;\ \theta\right) + w_k, & k &= 0, \dots, N-1,\\
y_k &= g(x_k; \theta) + v_k, & k &= 0, \dots, N.
\end{aligned}
$$

- The functions f and g are known, but it depends on unknown parameters θ.
- The variables x_k are the state of the system (usually unknown), while y_k are measurements (known). The initial state x_0 is in general, also unknown.
- The variables u_k are the inputs of the system (known). They can come from feedbacks from past measurements $y_0, \dots, y_{k-1}$.
- The random variables w_k and v_k are called the process noise and the measurement noise respectively. They are assumed to be independent with each other, and with zero-mean:

$$
\mathbb{E}\left[w_k\right] = \mathbb{E}\left[v_k\right] = 0
$$

Parametric Models

Parametric Dynamical Models with Random Noise

$$
\begin{aligned}
x_{k+1} &= f\big(x_k,\ u_k;\ \theta\big) + w_k, \qquad && k = 0, \ldots, N-1, \\
y_k &= g(x_k; \theta) + v_k, && k = 0, \ldots, N.
\end{aligned}
$$

- The functions f and g are known, but it depends on unknown parameters θ.
- The variables x_k are the state of the system (usually unknown), while y_k are measurements (known). The initial state x_0 is in general, also unknown.
- The variables u_k are the inputs of the system (known). They can come from feedbacks from past measurements $y_0, \ldots, y_{k-1}$.
- The random variables w_k and v_k are called the process noise and the measurement noise respectively. They are assumed to be independent with each other, and with zero-mean:

$$
\mathbb{E}\big[w_k\big] = \mathbb{E}\big[v_k\big] = 0
$$

$\Rightarrow$ **Goal:** estimate the parameters $\theta \in \mathbb{R}^{n_\theta}$ from data data $u_0, y_0, \ldots, u_{N-1}, y_N$.

Parametric Models

Parametric Linear Dynamical Model with Random Noise

Dynamical model:

$$
\begin{aligned}
x_{k+1} &= A\big(u_k;\theta\big)x_k + b\big(u_k;\theta\big) + w_k, && k = 0,\dots,N-1,\\
y_k &= C(\theta)x_k + v_k, && k = 0,\dots,N,
\end{aligned}
$$

Probabilistic model:

$$
\begin{aligned}
\mathrm{Cov}\big[w_k\big] &= Q\big(u_k;\theta\big), && k = 0,\dots,N-1,\\
\mathrm{Cov}\big[v_k\big] &= R\big(\theta\big), && k = 0,\dots,N-1,\\
\mathbb{E}\big[x_0\big] &= \hat{x}_0,\\
\mathrm{Cov}\big[x_0\big] &= P_0.
\end{aligned}
$$

Parametric Models

Parametric Linear Dynamical Model with Random Noise

Dynamical model:

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

Probabilistic model:

$$
\begin{aligned}
\mathrm{Cov}[w_k] &= Q(u_k;\theta), && k = 0,\dots,N-1,\\
\mathrm{Cov}[v_k] &= R(\theta), && k = 0,\dots,N-1,\\
\mathbb{E}[x_0] &= \hat{x}_0,\\
\mathrm{Cov}[x_0] &= P_0.
\end{aligned}
$$

- The functions $A(\cdot)$, $b(\cdot)$ and $C(\cdot)$ are known functions that parameterize the model.
- The functions $Q(\cdot)$ and $R(\cdot)$ parameterize the uncertainty model.

Parametric Models

Parametric Linear Dynamical Model with Random Noise

Dynamical model:

$$
\begin{aligned}
x_{k+1} &= A\big(u_k;\theta\big)x_k + b\big(u_k;\theta\big) + w_k, && k = 0,\ldots,N-1,\\
y_k &= C(\theta)x_k + v_k, && k = 0,\ldots,N,
\end{aligned}
$$

Probabilistic model:

$$
\begin{aligned}
\mathrm{Cov}\big[w_k\big] &= Q\big(u_k;\theta\big), && k = 0,\ldots,N-1,\\
\mathrm{Cov}\big[v_k\big] &= R\big(\theta\big), && k = 0,\ldots,N-1,\\
\mathbb{E}\big[x_0\big] &= \hat{x}_0,\\
\mathrm{Cov}\big[x_0\big] &= P_0.
\end{aligned}
$$

- The functions $A(\cdot)$, $b(\cdot)$ and $C(\cdot)$ are known functions that parameterize the model.
- The functions $Q(\cdot)$ and $R(\cdot)$ parameterize the uncertainty model.
- The prior knowledge on θ is in the form $h(\theta) \leq 0$.

Two tutorial examples

Two tutorial examples: the random walk model

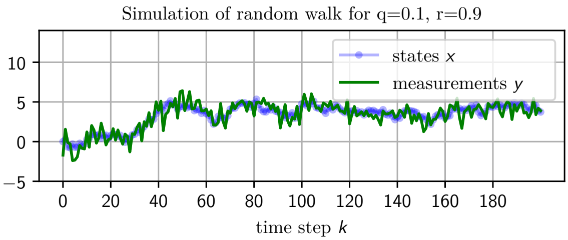
Random walk model with noisy measurements

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

Two tutorial examples: the random walk model

Random walk model with noisy measurements

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

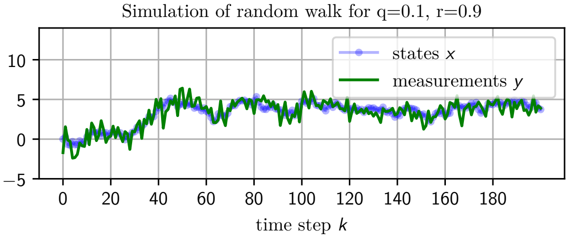


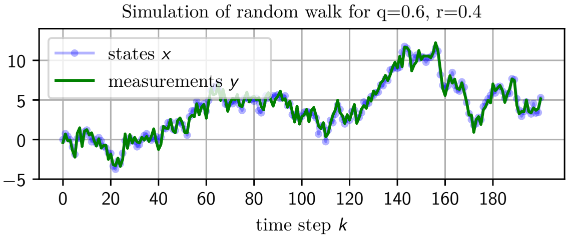
Two tutorial examples: the random walk model

Random walk model with noisy measurements

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

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





Two tutorial examples: a heat transfer model

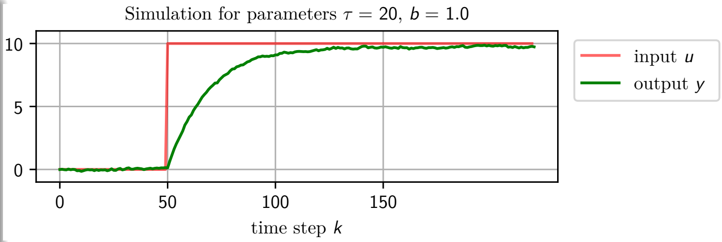
A heat transfer model

$$
\begin{aligned}
x_{k+1} &= (1 - 1/\tau)x_k + b/\tau\, u_k, \\
d_{k+1} &= d_k + w_k, \\
y_k &= x_k + d_k + v_k, \\
w_k &\sim \mathcal{N}(0, 10^{-3}), \\
v_k &\sim \mathcal{N}(0, 10^{-3}), \\
x_0 &= 0, \\
d_0 &= 0, \\
\theta &= \begin{bmatrix} 1/\tau & b \end{bmatrix}
\end{aligned}
$$

Two tutorial examples: a heat transfer model

A heat transfer model

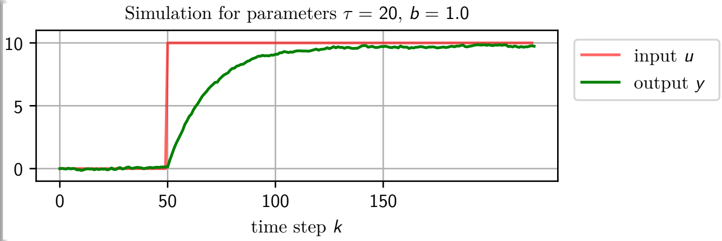
$$x_{k+1} = (1 - {}^1\!/\!_\tau)x_k + {}^b\!/\!_\tau u_k,$$
$$d_{k+1} = d_k + w_k,$$
$$y_k = x_k + d_k + v_k,$$
$$w_k \sim \mathcal{N}(0, 10^{-3}),$$
$$v_k \sim \mathcal{N}(0, 10^{-3}),$$
$$x_0 = 0,$$
$$d_0 = 0,$$
$$\theta = \begin{bmatrix} {}^1\!/\!_\tau & b \end{bmatrix}$$

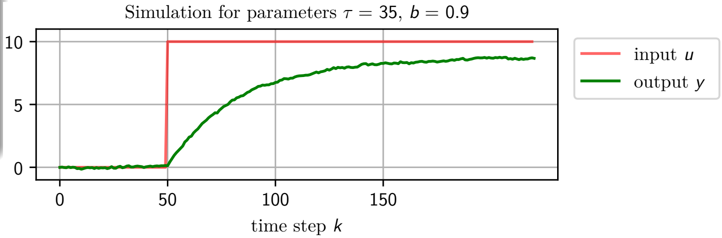


Two tutorial examples: a heat transfer model

A heat transfer model

$$
\begin{aligned}
x_{k+1} &= (1 - 1/\tau)x_k + b/\tau\, u_k, \\
d_{k+1} &= d_k + w_k, \\
y_k &= x_k + d_k + v_k, \\
w_k &\sim \mathcal{N}(0, 10^{-3}), \\
v_k &\sim \mathcal{N}(0, 10^{-3}), \\
x_0 &= 0, \\
d_0 &= 0, \\
\theta &= \begin{bmatrix} 1/\tau & b \end{bmatrix}
\end{aligned}
$$





Is the noise model that important?

Is the noise model that important?

$$
\begin{aligned}
x_{k+1} &= f(x_k, u_k; \theta) + w_k, & k &= 0, \ldots, N-1, \\
y_k &= g(x_k, u_k; \theta) + v_k, & k &= 0, \ldots, N,
\end{aligned}
$$

As we will see in the next part, estimating a noise model is useful for the design of a good state estimator.
But what if one is only interested by estimating a nominal model?
Could we ignore the noise, and estimate a deterministic model?

Is the noise model that important?

$$
\begin{aligned}
x_{k+1} &= f\left(x_k, u_k; \theta\right) + w_k, & k &= 0, \ldots, N-1, \\
y_k &= g(x_k, u_k; \theta) + v_k, & k &= 0, \ldots, N,
\end{aligned}
$$

Could we ignore the noise, and estimate a deterministic model?

A positive answer: There exists methods that provide estimates $\hat{\theta}_N$ that are robust against the noise given some (reasonable) assumptions

Robust estimation

In robust estimation, we look for an estimate $\hat{\theta}_N$ that converges to the true value $\theta^\star$ if the process and measurement noise goes to zero when $k \to +\infty$. A sufficient condition:

$$
\left\| \hat{\theta}_N - \theta^\star \right\| \leq C_1 \rho_1^N + C_2 \max_{k=1,\ldots,N} \rho_2^{N-k} \|w_k\| + C_3 \max_{k=1,\ldots,N} \rho_3^{N-k} \|v_k\|,
$$

... but can't we do better by using $\mathbb{E}[w_k] = \mathbb{E}[v_k] = 0$?

A simple instructive example

Let us have a look at one simple example:

A simple example

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

One greedy idea to estimate θ:

Since we have a (noisy) measurement of x_k, we could just replace it with y_k on the first equation, then do linear regression to find θ...

$$\begin{aligned} y_{k+1} &= \theta\left(y_k - v_k\right) + w_k + v_{k+1}, \\ &= \theta y_k + w_k + v_{k+1} - \theta v_k, \end{aligned}$$

$$\overbrace{y_{k+1}}^{\text{known}} = \theta \overbrace{y_k}^{\text{known}} + \overbrace{w_k + v_{k+1} - \theta v_k}^{\text{noise}},$$

Linear regression to find θ:

$$\hat{\theta} = \frac{\sum_{k=1}^{N} y_{k+1} y_k}{\sum_{k=1}^{N} {y_k}^2}$$

A simple instructive example

$$\overbrace{y_{k+1}}^{\text{known}} = \theta \overbrace{y_k}^{\text{known}} + \overbrace{w_k + v_{k+1} - \theta v_k}^{\text{noise}},$$

Linear regression to find θ:

$$\begin{aligned}
\hat{\theta}_N &= \frac{\sum_{k=1}^{N} y_{k+1} y_k}{\sum_{k=1}^{N} {y_k}^2}, \\
&= \frac{\sum_{k=1}^{N} \theta y_k y_k}{\sum_{k=1}^{N} {y_k}^2} + \frac{\sum_{k=1}^{N} \left[w_k + v_{k+1} - \theta v_k \right] y_k}{\sum_{k=1}^{N} {y_k}^2}, \\
&= \theta + \frac{\frac{1}{N} \sum_{k=1}^{N} \left[w_k + v_{k+1} - \theta v_k \right] y_k}{\frac{1}{N} \sum_{k=1}^{N} {y_k}^2}, \\
&\xrightarrow[N \to \infty]{} \theta + \frac{\mathbb{E}\left[w_k y_k \right]}{\mathbb{E}\left[{y_k}^2 \right]} + \frac{\mathbb{E}\left[v_{k+1} y_k \right]}{\mathbb{E}\left[{y_k}^2 \right]} - \theta \frac{\mathbb{E}\left[v_k y_k \right]}{\mathbb{E}\left[{y_k}^2 \right]} = \theta - \theta \frac{\mathbb{E}\left[{v_k}^2 \right]}{\mathbb{E}\left[{y_k}^2 \right]}
\end{aligned}$$

A simple example

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

Conclusion:

If one tries to estimate parameters ”as if the measurements weren't noisy”, the estimate would be biased, even for infinite amount of data:

$$\hat{\theta}_N \underset{N\to\infty}{\longrightarrow} \theta\left(1 - \frac{\mathbb{E}\left[{v_k}^2\right]}{\mathbb{E}\left[{y_k}^2\right]}\right)$$

The Kalman Filter

Linear Time Variant system

Parametric Linear Dynamical Model with Random Noise

Dynamical model:

$$\begin{aligned} x_{k+1} &= A\big(u_k;\theta\big)x_k + b\big(u_k;\theta\big) + w_k, && k = 0,\dots,N-1,\\ y_k &= C(\theta)x_k + v_k, && k = 0,\dots,N, \end{aligned}$$

Probabilistic model:

$$\begin{aligned} \mathrm{Cov}\big[w_k\big] &= Q\big(u_k;\theta\big), && k = 0,\dots,N-1,\\ \mathrm{Cov}\big[v_k\big] &= R\big(\theta\big), && k = 0,\dots,N-1,\\ \mathbb{E}\big[x_0\big] &= \hat{x}_0,\\ \mathrm{Cov}\big[x_0\big] &= P_0. \end{aligned}$$

Linear Time Variant system

In this part, we omit the dependency in θ and u_k, and focuses on the estimation of the state x_k:

Linear Time-Variant System with Random Noise

Dynamical model:

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

Probabilistic model:

$$\begin{aligned} \mathrm{Cov}\left[w_k\right] &= Q_k, && k = 0, \dots, N-1, \\ \mathrm{Cov}\left[v_k\right] &= R, && k = 0, \dots, N-1, \\ \mathbb{E}\left[x_0\right] &= \hat{x}_0, \\ \mathrm{Cov}\left[x_0\right] &= P_0. \end{aligned}$$

The Kalman Filter

Dynamical model:

$$x_{k+1} = A_k x_k + b_k + w_k,$$
$$y_k = C x_k + v_k,$$

Probabilistic model:

$$\mathrm{Cov}\left[w_k\right] = Q_k,$$
$$\mathrm{Cov}\left[v_k\right] = R,$$
$$\mathbb{E}\left[x_0\right] = \hat{x}_0,$$
$$\mathrm{Cov}\left[x_0\right] = P_0.$$

- We initialize our estimate with $\hat{x}_0 = \hat{x}_0$
- Everytime a new measurement y_k is available, we update our state estimate:

$$e_k = y_k - C\,\hat{x}_{k|k-1},$$
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k e_k$$

- Then we predict the state at the next time step

$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + b_k,$$

... But how to choose K_k ?

The Kalman Filter

Dynamical model:

$$\begin{aligned} x_{k+1} &= A_k x_k + b_k + w_k, \\ y_k &= C x_k + v_k, \end{aligned}$$

Probabilistic model:

$$\begin{aligned} \mathrm{Cov}\left[w_k\right] &= Q_k, \\ \mathrm{Cov}\left[v_k\right] &= R, \\ \mathbb{E}\left[x_0\right] &= \hat{x}_0, \\ \mathrm{Cov}\left[x_0\right] &= P_0, \end{aligned}$$

The Kalman Filter:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\left(y_k - C\,\hat{x}_{k|k-1}\right), \\ \hat{x}_{k+1|k} &= A_k \hat{x}_{k|k} + b_k. \end{aligned}$$

Define

$$\begin{aligned} P_{k|k} &:= \mathrm{Cov}\left[x_k - \hat{x}_{k|k}\right], \\ P_{k|k-1} &:= \mathrm{Cov}\left[x_k - \hat{x}_{k|k-1}\right]. \end{aligned}$$

Then the following equations hold:

$$\begin{aligned} P_{k|k} &= (I - K_k C) P_{k|k-1} (I - K_k C)^\top + K_k R {K_k}^\top, \\ P_{k+1|k} &= A_k P_{k|k} A_k^\top + Q_k, \end{aligned}$$

$\Rightarrow$ We chose $K_k = P_{k|k-1} C^\top \left(C P_{k|k-1} C^\top + R\right)^{-1}$ because it minimizes $P_{k|k}$ in every direction. (proof on the blackboard)

The Kalman Filter Equations

To summarize, the Kalman Filter is given by the following equations

Dynamical model:

$$x_{k+1} = A_k x_k + b_k + w_k,$$
$$y_k = C x_k + v_k,$$

Probabilistic model:

$$\mathrm{Cov}\left[w_k\right] = Q_k,$$
$$\mathrm{Cov}\left[v_k\right] = R,$$
$$\mathbb{E}\left[x_0\right] = \hat{x}_0,$$
$$\mathrm{Cov}\left[x_0\right] = P_0,$$

The Kalman Filter

- **Initialize** the state and covariance estimates with $\hat{x}_0$ and P_0
- Compute the Kalman gain
$$K_k = P_{k|k-1} C^\top \left(C P_{k|k-1} C^\top + R\right)^{-1}$$
- **Update** with respect to new measurement y_k:
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k\left(y_k - C\,\hat{x}_{k|k-1}\right),$$
$$P_{k|k} = (I - K_k C) P_{k|k-1} (I - K_k C)^\top + K_k R K_k{}^\top,$$
- **Predict** the next state:
$$\hat{x}_{k+1|k} = A_k \hat{x}_{k|k} + b_k,$$
$$P_{k+1|k} = A_k P_{k|k} A_k^\top + Q_k.$$

The Kalman Filter Equations

There are many other equivalent ways of formulating the Kalman Filter equations. Here is another, more compact:

The Kalman FIlter Equations (more compact)

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

In the case of Gaussian noise :

Dynamical model:

$$x_{k+1} = A_k x_k + b_k + w_k,$$
$$y_k = C x_k + v_k,$$

Probabilistic model:

$$w_k \sim \mathcal{N}(0_{n_x}, Q_k),$$
$$v_k \sim \mathcal{N}(0_{n_y}, R),$$
$$x_0 \sim \mathcal{N}(\hat{x}_0, P_0).$$

The following properties hold

$$\left(x_k \mid y_0, \ldots, y_{k-1}\right) \sim \mathcal{N}\left(\hat{x}_{k|k-1}, P_{k|k-1}\right),$$
$$\left(x_k \mid y_0, \ldots, y_k\right) \sim \mathcal{N}\left(\hat{x}_{k|k}, P_{k|k}\right)$$

We have also a similar property on the output y_k:

$$\left(y_k \mid y_0, \ldots, y_{k-1}\right) \sim \mathcal{N}\left(C\hat{x}_{k|k-1}, S_k\right),$$

with S_k *innovation covariance*, defined as follows

$$S_k := C P_{k|k-1} C^\top + R$$

Prediction Error Method: an appropriate method for our problem

Parametric Models

Let us recall our initial problem:

Parametric Linear Dynamical Model with Random Noise

Dynamical model:

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

Probabilistic model:

$$\begin{aligned} \mathrm{Cov}\left[w_k\right] &= Q(u_k;\theta), & k &= 0,\dots,N-1, \\ \mathrm{Cov}\left[v_k\right] &= R(\theta), & k &= 0,\dots,N-1, \\ \mathbb{E}\left[x_0\right] &= \hat{x}_0, \\ \mathrm{Cov}\left[x_0\right] &= P_0. \end{aligned}$$

Kalman Filter on our problem

The equations of the Kalman Filter

For given parameters θ, the Kalman Filter is given by the following equations:

$$
\begin{aligned}
S_k &= C P_k C^\top + R(\theta), & k &= 0, \dots, N, \\
L_k &= A(u_k;\theta) P_k C^\top {S_k}^{-1}, & k &= 0, \dots, N, \\
\hat{x}_{k+1} &= \big(A(u_k;\theta) - L_k C\big)\hat{x}_k + L_k y_k + b(u_k;\theta), & k &= 0, \dots, N-1, \\
P_{k+1} &= A(u_k;\theta) P_k A(u_k;\theta)^\top - L_k S_k {L_k}^\top + Q(u_k;\theta), & k &= 0, \dots, N-1.
\end{aligned}
$$

Kalman Filter on our problem

The equations of the Kalman Filter

For given parameters θ, the Kalman Filter is given by the following equations:

$$
\begin{aligned}
S_k &= C P_k C^\top + R(\theta), && k = 0, \dots, N,\\
L_k &= A\big(u_k;\theta\big) P_k C^\top {S_k}^{-1}, && k = 0, \dots, N,\\
\hat{x}_{k+1} &= \big(A\big(u_k;\theta\big) - L_k C\big)\hat{x}_k + L_k y_k + b\big(u_k;\theta\big), && k = 0, \dots, N-1,\\
P_{k+1} &= A\big(u_k;\theta\big) P_k A\big(u_k;\theta\big)^\top - L_k S_k {L_k}^\top + Q\big(u_k;\theta\big), && k = 0, \dots, N-1.
\end{aligned}
$$

We define the functions "$\hat{y}_k(y_0, \dots, y_{k-1}, \theta) := C\hat{x}_k$" and "$S_k(\theta) := S_k$".

Kalman Filter on our problem

The equations of the Kalman Filter

For given parameters θ, the Kalman Filter is given by the following equations:

$$
\begin{aligned}
S_k &= C P_k C^\top + R(\theta), && k = 0, \dots, N, \\
L_k &= A\big(u_k; \theta\big) P_k C^\top S_k{}^{-1}, && k = 0, \dots, N, \\
\hat{x}_{k+1} &= \big(A\big(u_k; \theta\big) - L_k C\big)\hat{x}_k + L_k y_k + b\big(u_k; \theta\big), && k = 0, \dots, N-1, \\
P_{k+1} &= A\big(u_k; \theta\big) P_k A\big(u_k; \theta\big)^\top - L_k S_k L_k{}^\top + Q\big(u_k; \theta\big), && k = 0, \dots, N-1.
\end{aligned}
$$

We define the functions "$\hat{y}_k(y_0, \dots, y_{k-1}, \theta) := C\hat{x}_k$" and "$S_k(\theta) := S_k$".

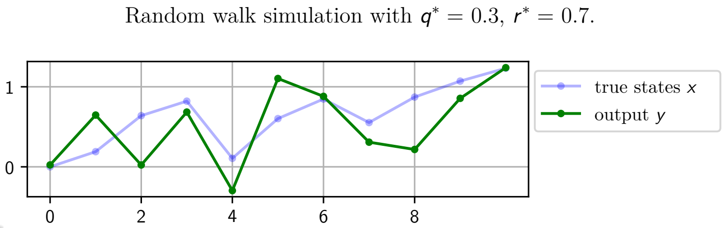
$\Rightarrow$ **When the noise is Gaussian, the following holds:**

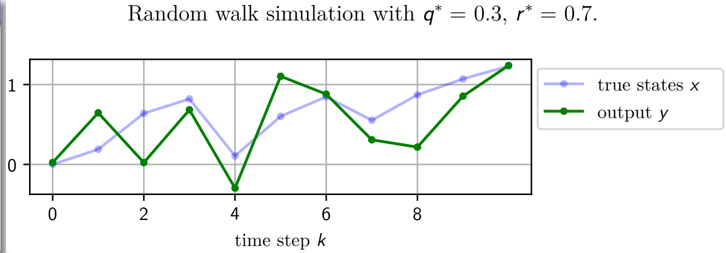
$$
\big(y_k \mid y_0, \dots, y_{k-1}, \theta\big) \sim \mathcal{N}\Big(\hat{y}_k\big(y_0, \dots, y_{k-1}, \theta\big), \quad S_k(\theta)\Big)
$$

- 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

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

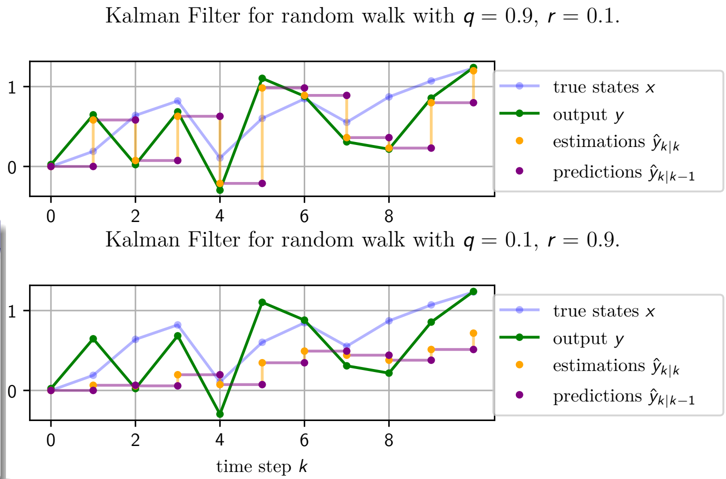




- We generate data with the random walk model, with covariances $q^* = 0.3$ and $r^* = 0.7$
- We apply a KF with other values of q and r.

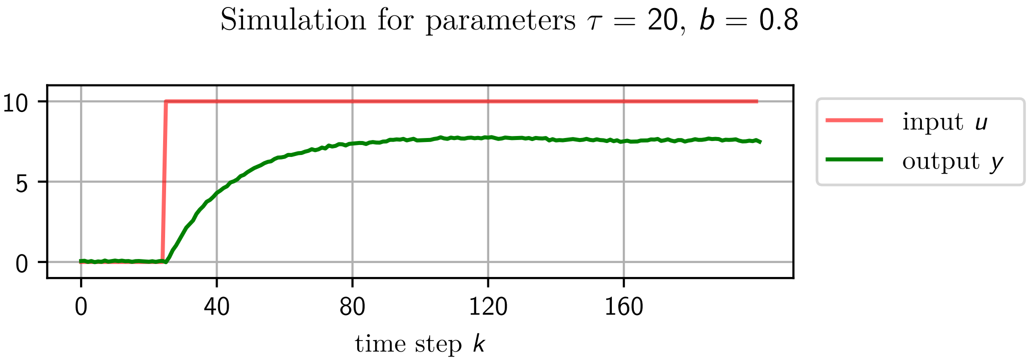
Random walk model

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

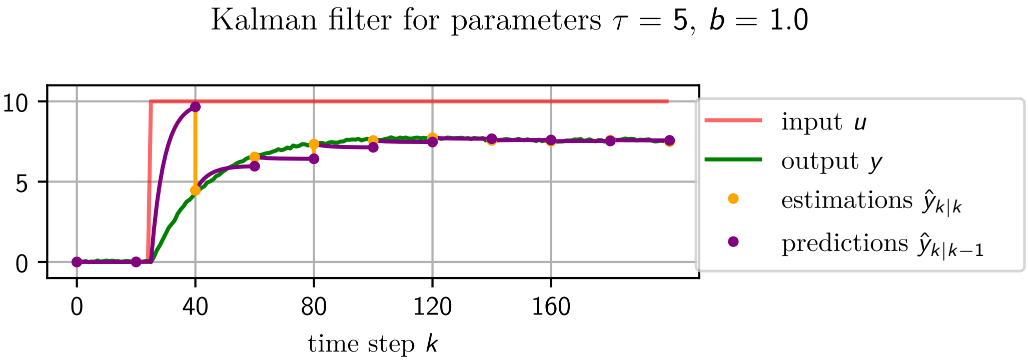


The Kalman Filter in the heat transfer model

- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.

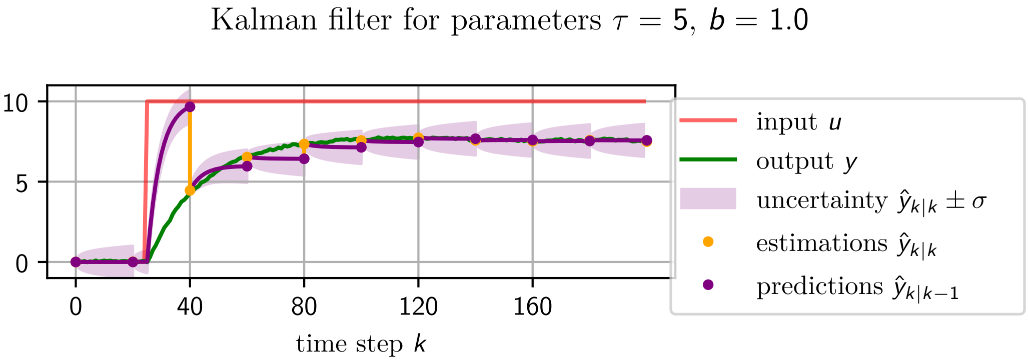


- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.



The Kalman Filter in the heat transfer model

- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.



Qualitative description

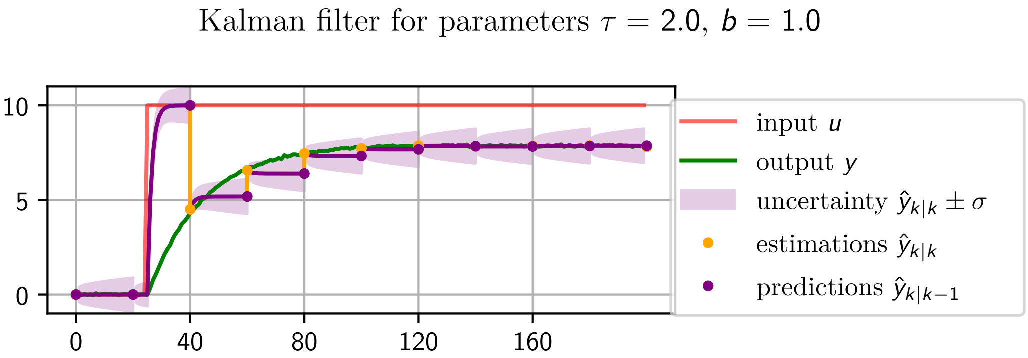
- One can apply a Kalman Filter to the measurement data for given parameters θ.
- We seek for the parameters resulting in the "best KF".
- We measure the quality of the KF with the prediction error $y_k - \hat{y}_k(y_0, \ldots, y_{k-1}, \theta)$.
- This is refined by considering not only the prediction error, but also its estimated covariance $S_k(\theta)$.
- This method belongs to the class of *prediction error estimation methods* [1], and it is a *Maximum Likelihood Estimation* method. [2].

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

[2] R. Kashyap, "Maximum likelihood identification of stochastic linear systems," *IEEE Transactions on Automatic Control, vol. 15, no. 1, pp.25–34*, 1970

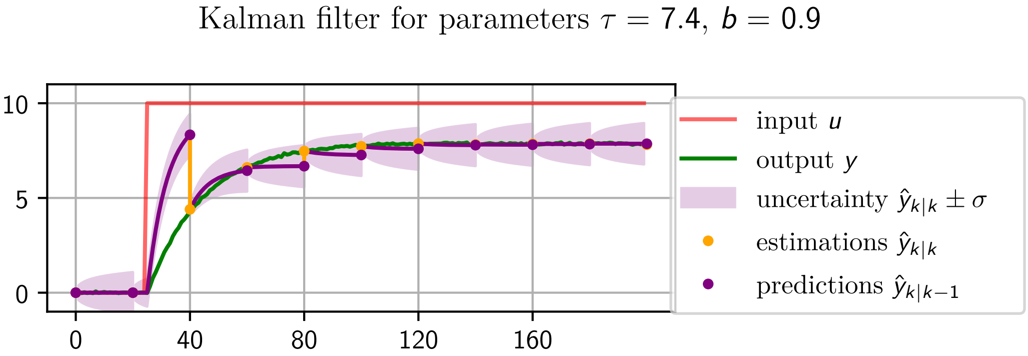
Some optimization steps in the heat transfer model

- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.

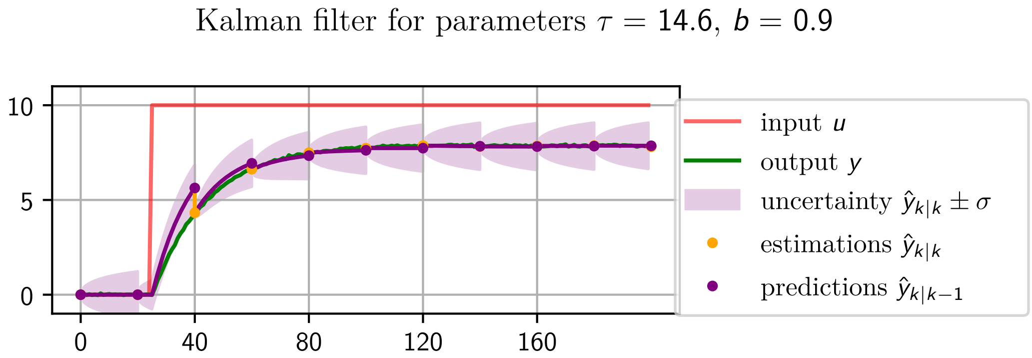


Some optimization steps in the heat transfer model

- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.



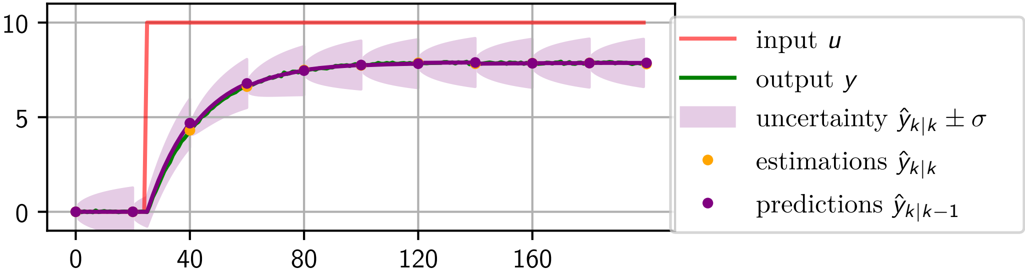
- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.



Some optimization steps in the heat transfer model

- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.

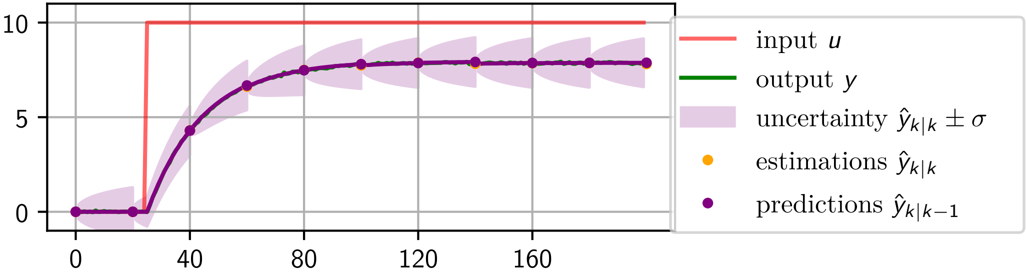
Kalman filter for parameters $\tau = 18.2$, $b = 0.8$



Some optimization steps in the heat transfer model

- We generate data with the heat transfer model, with parameters $\tau^* = 20$ and $b^* = 0.8$
- We apply a KF with other values of τ and b.

Kalman filter for parameters $\tau = 20.0$, $b = 0.8$



The optimization problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \sum_{k=0}^{N} \left\| y_k - \hat{y}_k\left(y_0, \ldots, y_{k-1}, \theta\right) \right\|^2_{S_k(\theta)^{-1}} + \log \left| S_k(\theta) \right| \\ \text{subject to} \quad & h(\theta) \leq 0. \end{aligned}$$

The optimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{k=0}^{N} \left\| y_k - \hat{y}_k\left(y_0, \ldots, y_{k-1}, \theta\right) \right\|^2_{S_k(\theta)^{-1}} + \log\left|S_k(\theta)\right|$$

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

Notations:

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

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

The optimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{k=0}^{N} \left\| y_k - \hat{y}_k\left(y_0, \dots, y_{k-1}, \theta\right) \right\|^2_{S_k(\theta)^{-1}} + \log \left| S_k(\theta) \right|$$

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

Notations:

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

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

Lifted form:

$$\underset{\theta, \boldsymbol{S}, \boldsymbol{L}, \boldsymbol{e}, \boldsymbol{x}, \boldsymbol{P}}{\text{minimize}} \quad \sum_{k=0}^{N} \|e_k\|^2_{S_k^{-1}} + \log |S_k|$$

$$\begin{aligned}
\text{subject to} \quad & S_k = C P_k C^\top + R(\theta), && k = 0, \dots, N, \\
& L_k = A\left(u_k; \theta\right) P_k C^\top S_k{}^{-1}, && k = 0, \dots, N, \\
& e_k = y_k - C\hat{x}_k && k = 0, \dots, N, \\
& \hat{x}_{k+1} = A\left(u_k; \theta\right)\hat{x}_k + L_k e_k + b\left(u_k; \theta\right), && k = 0, \dots, N-1, \\
& P_{k+1} = A\left(u_k; \theta\right) P_k A\left(u_k; \theta\right)^\top - L_k S_k L_k{}^\top + Q(u_k; \theta), && k = 0, \dots, N-1, \\
& h(\theta) \leq 0.
\end{aligned}$$

Stationary Kalman Filter for Liner Time-Invariant Systems

Linear Time-Invariant Systems

Dynamical model:

$$x_{k+1} = A(\theta)x_k + b(u_k;\theta) + w_k,$$
$$y_k = C(\theta)x_k + v_k,$$

Probabilistic model:

$$\mathrm{Cov}[w_k] = Q(\theta),$$
$$\mathrm{Cov}[v_k] = R(\theta),$$
$$\mathbb{E}[x_0] = \hat{x}_0,$$
$$\mathrm{Cov}[x_0] = P_0.$$

In the present case, the Kalman Filter equations would converge (exponentially) to its steady state (aka Ricatti Equation):

$$
\begin{aligned}
e_k &= y_k - C\hat{x}_k, && k = 0, \dots \\
\hat{x}_{k+1} &= A(\theta)\hat{x}_k + Le_k + b(u_k;\theta), && k = 0, \dots \\
S &= CPC^\top + R(\theta), \\
L &= A(\theta)PC^\top S^{-1}, \\
P &= A(\theta)PA(\theta)^\top - LSL^\top + Q(\theta),
\end{aligned}
$$

Now also the optimization problem also simplifies a lot:

The optimization problem for LTI systems

$$
\begin{aligned}
\underset{\theta, S, L, P, \boldsymbol{x}}{\text{minimize}} \quad & \log|S| + \mathrm{Tr}\left(S^{-1}\frac{1}{N}\sum_{k=1}^{N} e_k {e_k}^\top\right) & \\
\text{subject to} \quad & e_k = y_k - C\hat{x}_k, & k = 0, \dots, N \\
& \hat{x}_{k+1} = A(\theta)\hat{x}_k + Le_k + b\left(u_k; \theta\right), & k = 0, \dots, N-1 \\
& S = CPC^\top + R(\theta), & \\
& L = A(\theta)PC^\top S^{-1}, & \\
& P = A(\theta)PA(\theta)^\top - LSL^\top + Q(\theta), & \\
& h(\theta) \leq 0. &
\end{aligned}
$$

Optimizing over L and S directly

When Q and R are fully unknown, it might be more convenient to optimize over L and S:

$$\begin{aligned} \underset{\theta, S, L, \mathbf{x}}{\text{minimize}} \quad & \log|S| + \operatorname{Tr}\left(S^{-1}\frac{1}{N}\sum_{k=1}^{N} e_k {e_k}^\top\right) \\ \text{subject to} \quad & e_k = y_k - C\hat{x}_k, && k = 0, \dots, N \\ & \hat{x}_{k+1} = A(\theta)\hat{x}_k + L e_k + b(u_k;\theta), && k = 0, \dots, N-1 \\ & h(\theta) \leq 0. \end{aligned}$$

The optimization over S can be done analytically: $S = \frac{1}{N}\sum_{k=1}^{N} e_k {e_k}^\top$,

$\Rightarrow$

$$\begin{aligned} \underset{\theta, L, \mathbf{x}}{\text{minimize}} \quad & \det\left(\frac{1}{N}\sum_{k=1}^{N} e_k {e_k}^\top\right) \\ \text{subject to} \quad & e_k = y_k - C\hat{x}_k, && k = 0, \dots, N \\ & \hat{x}_{k+1} = A(\theta)\hat{x}_k + L e_k + b(u_k;\theta), && k = 0, \dots, N-1 \\ & h(\theta) \leq 0. \end{aligned}$$

Theory about this estimation method

Theory about this estimation method

In this section, we will define

$$V_N(\boldsymbol{y}_N, \theta) = V_N(y_0, \ldots, y_N, \theta) = \frac{1}{N} \sum_{k=1}^{N} \left\| y_k - \hat{y}_k \left(y_0, \ldots, y_{k-1}, \theta \right) \right\|^2_{S_k(\theta)^{-1}} + \log |S_k(\theta)|$$

Then, our estimation method can be stated as follows:

$$\hat{\theta}_N(\boldsymbol{y}_N) = \arg\min_{\theta} V_N(\boldsymbol{y}_N, \theta)$$

Relation with Maximum Likelihood Estimation

Maximum Likelihood Estimation

Theorem: In the case of Gaussian noise, the former optimization problem is equivalent to the Maximum Likelihood Estimation:

$$V_N(y_0, \dots, y_N, \theta) = -2 \log \Big(p\left(y_0, \dots, y_N \mid \theta\right) \Big) + cst$$

Proof :

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

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

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

Expected value of the objective function

Assume that $y_0, \ldots, y_N$ follows the distribution of given by the same probabilistic model, with some parameters $\theta^\star$. Then $\theta^\star$ is a minimizer of the expected value of the objective function:

$$\theta^\star \in \arg\min_{\theta} \mathop{\mathbb{E}}_{y_0,\ldots,y_N} \left[V_N(y_0, \ldots, y_N, \theta) \quad \middle| \theta^\star \right].$$

Expected value of the objective function

Assume that $y_0, \ldots, y_N$ follows the distribution of given by the same probabilistic model, with some parameters $\theta^\star$. Then $\theta^\star$ is a minimizer of the expected value of the objective function:

$$\theta^\star \in \arg\min_{\theta} \mathop{\mathbb{E}}_{y_0,\ldots,y_N} \left[V_N(y_0, \ldots, y_N, \theta) \quad \middle| \theta^\star \right].$$

Proof : (in the Gaussian case) This property is inherited from the fact that the objective is the negative log likelihood of the measurements (up to multiplicative and additive constants). For some arbitrary parameters θ, let us define the probability densities $\hat{q}(\cdot) := p(\cdot \mid \theta)$ and $q^\star(\cdot) := p(\cdot \mid \theta^\star)$. One can show the desired property using Jensen's inequality on the convex function $-\log(\cdot)$:

$$\mathop{\mathbb{E}}_{\boldsymbol{y}} \left[-\log\left(p(\boldsymbol{y} \mid \theta)\right) \middle| \theta^\star \right] = \int_{\boldsymbol{y}} -\log\left(\hat{q}(\boldsymbol{y})\right) q^\star(\boldsymbol{y}) d\boldsymbol{y} \geq \int_{\boldsymbol{y}} -\log\left(q^\star(\boldsymbol{y})\right) q^\star(\boldsymbol{y}) d\boldsymbol{y}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{y}} \left[-\log\left(p(\boldsymbol{y} \mid \theta^\star)\right) \middle| \theta^\star \right].$$

Convergence of the estimate

Consistency of the estimate

Assume that $y_0, \ldots, y_N$ follows the distribution of given by the same probabilistic model, with some parameters $\theta^\star$. Let $\hat{\theta}_N(\boldsymbol{y})$ be the parameter estimate given by the presented method. Then, under some reasonable assumptions:

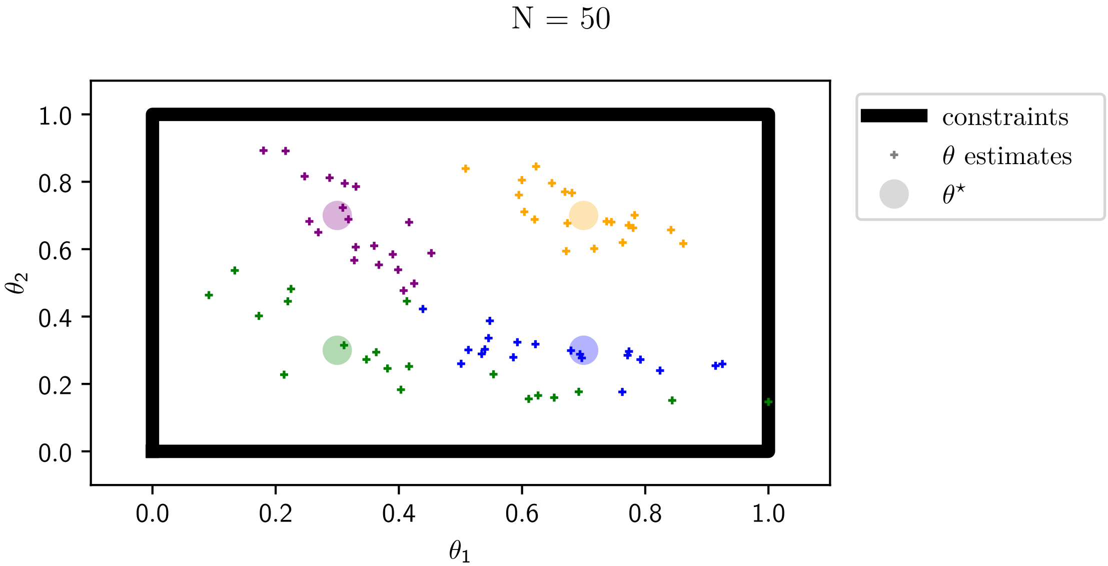
$$\hat{\theta}_N(\boldsymbol{y}) \underset{N \to +\infty}{\longrightarrow} \theta^\star \quad \text{with probability 1.}$$

Aymptotic convergence

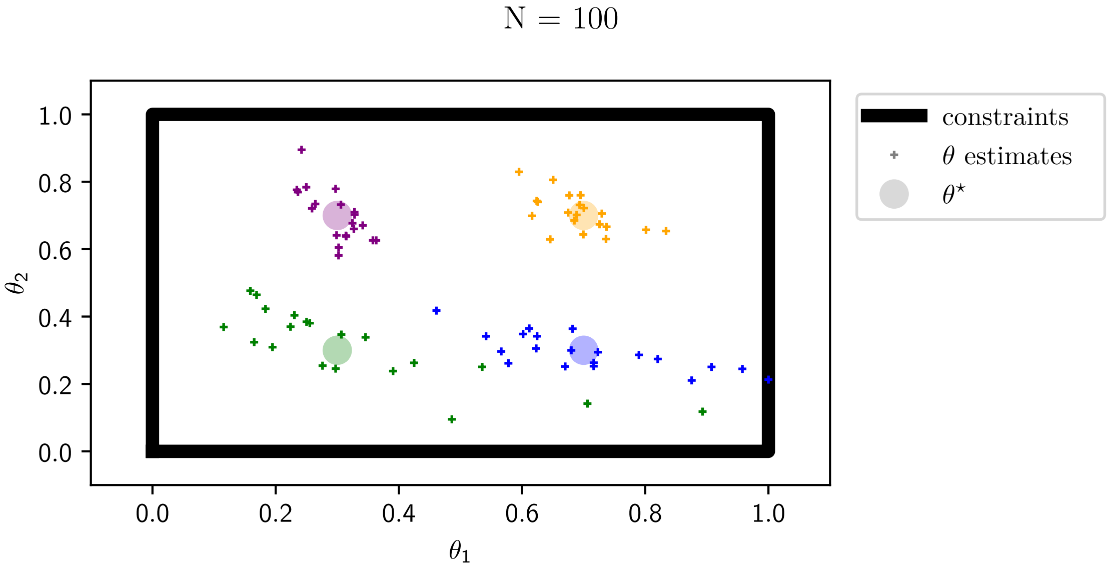
Furthermore, the following holds for some constant c:

$$\mathbb{E}_{\boldsymbol{y}}\left[\left\|\hat{\theta}_N(\boldsymbol{y}) - \theta^\star\right\|^2\right] = \frac{c}{N} + o\left(\frac{1}{N}\right)$$

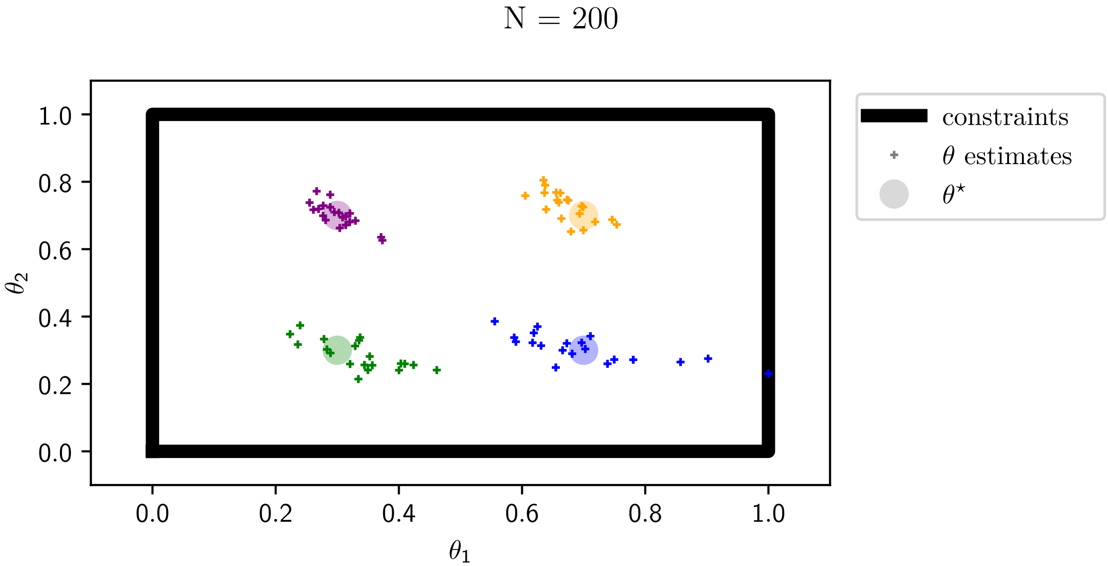
Performance on the heat transfer model



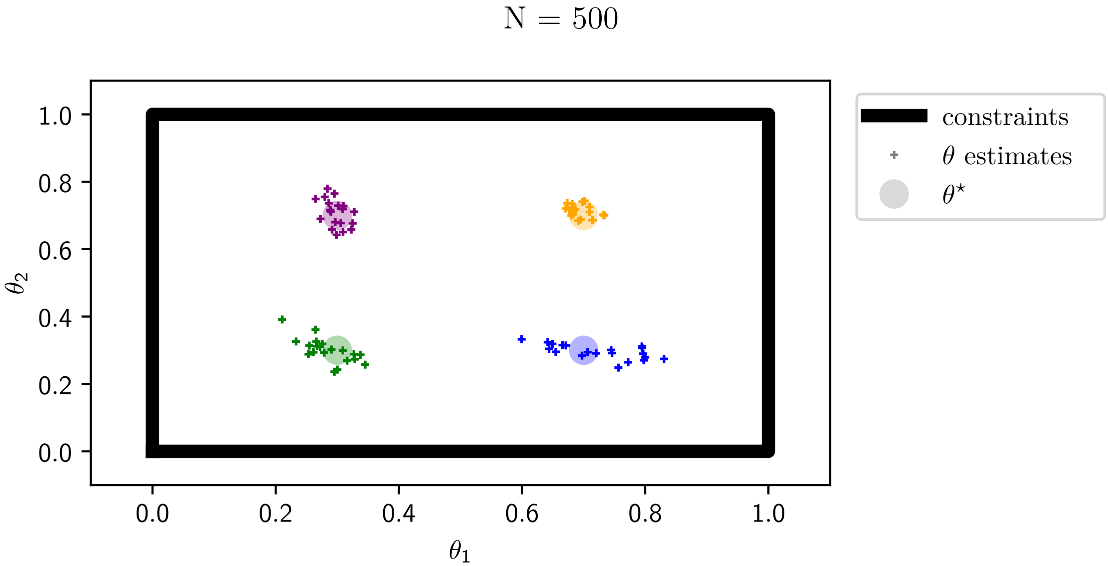
Performance on the heat transfer model



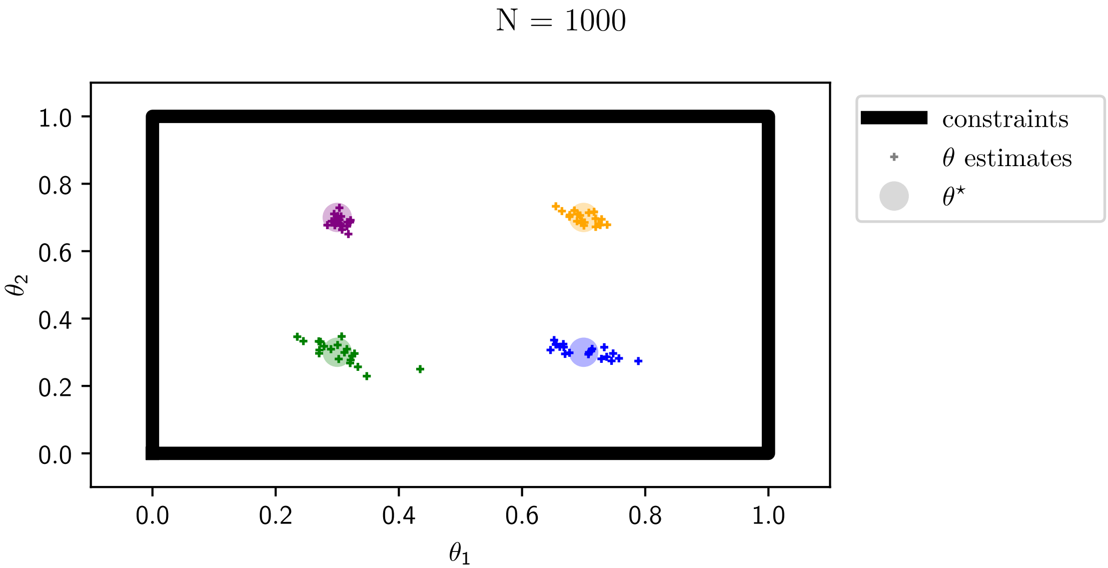
Performance on the heat transfer model



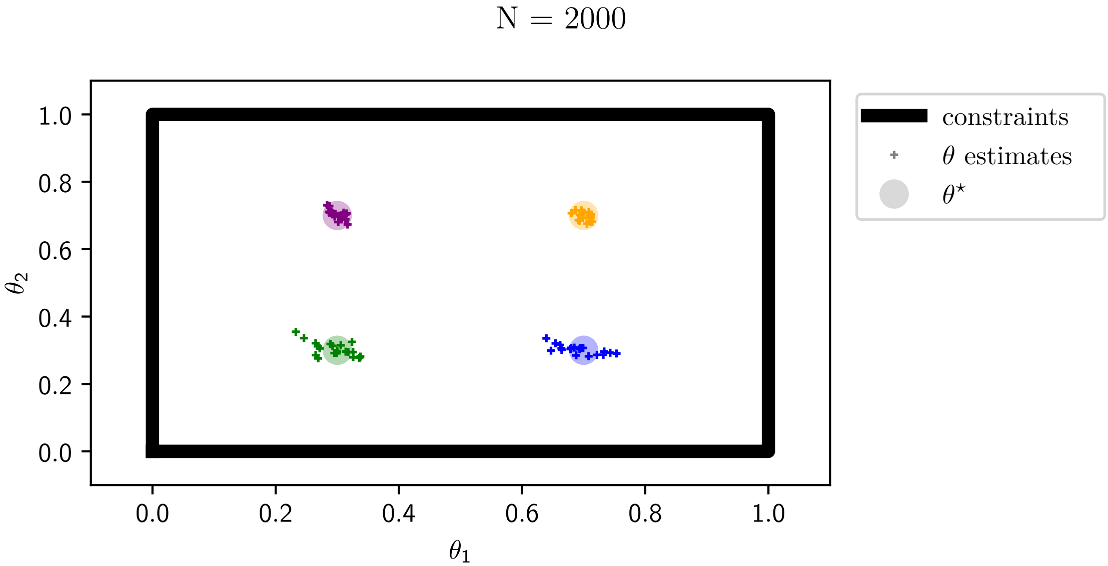
Performance on the heat transfer model



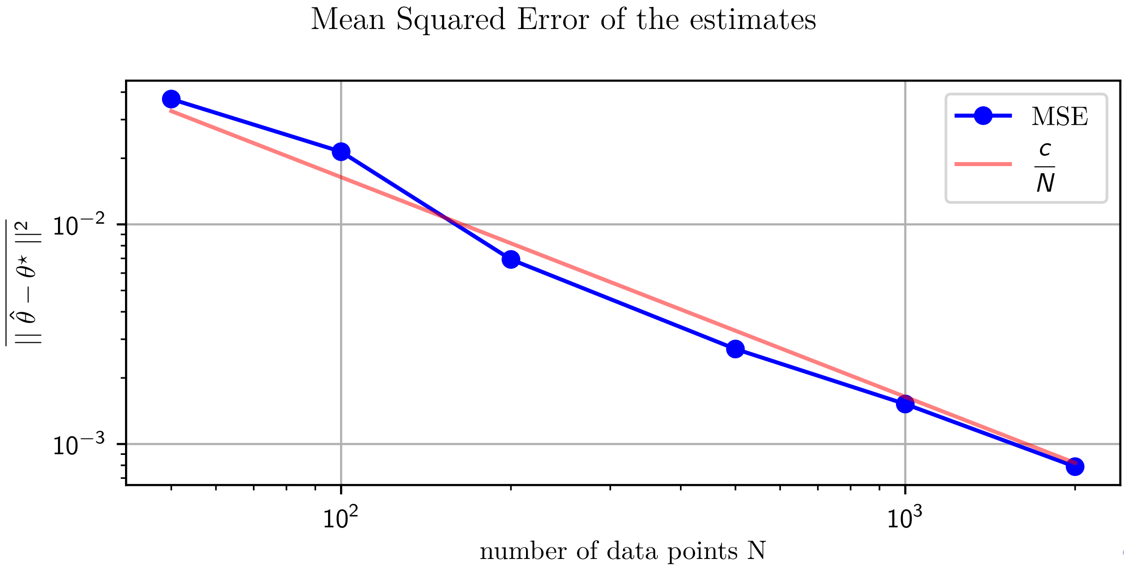
Performance on the heat transfer model



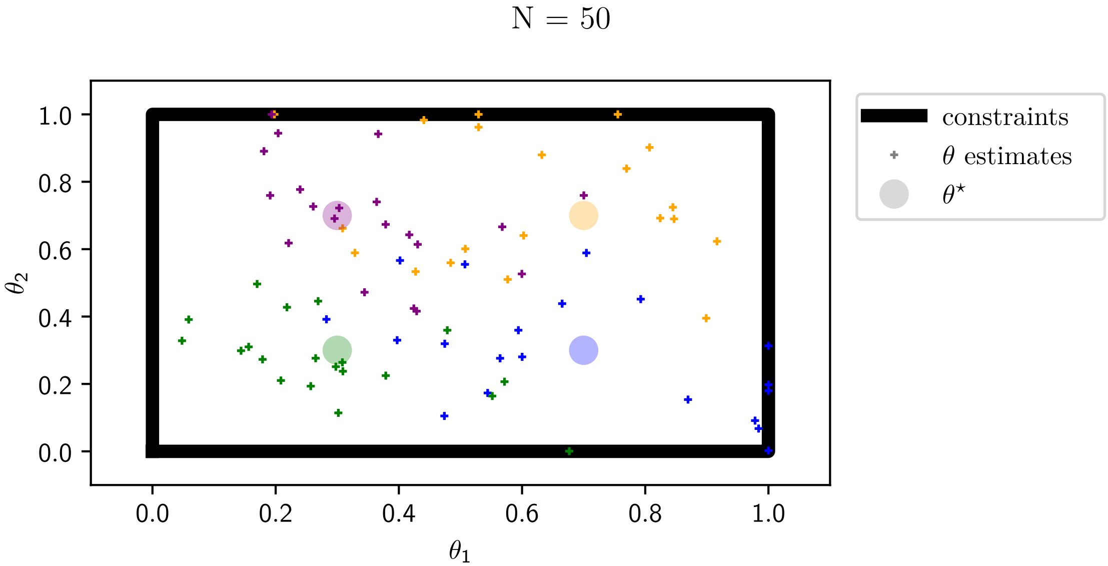
Performance on the heat transfer model

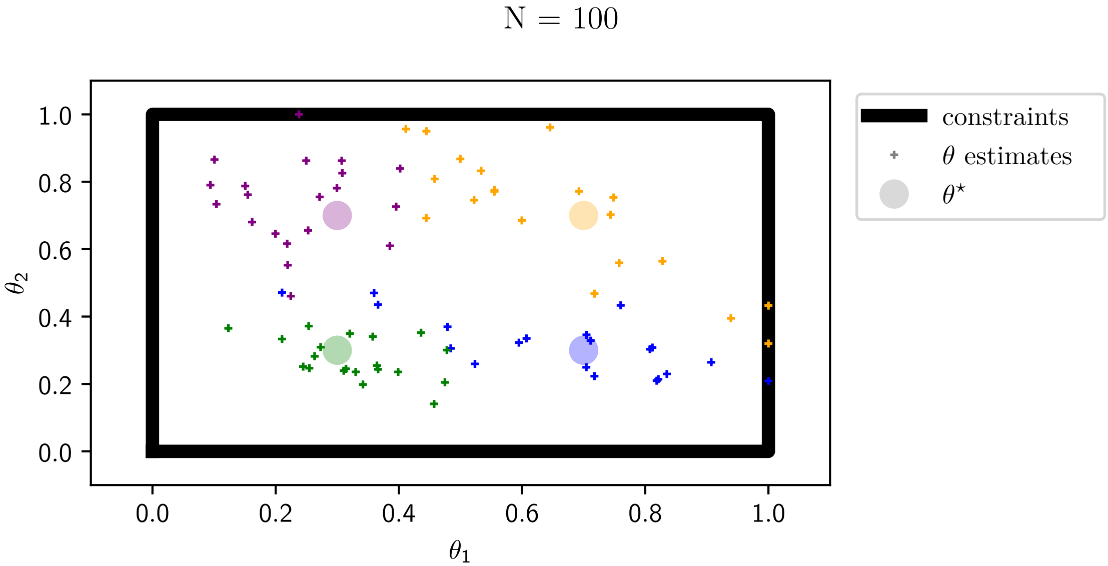


Performance on the heat transfer model

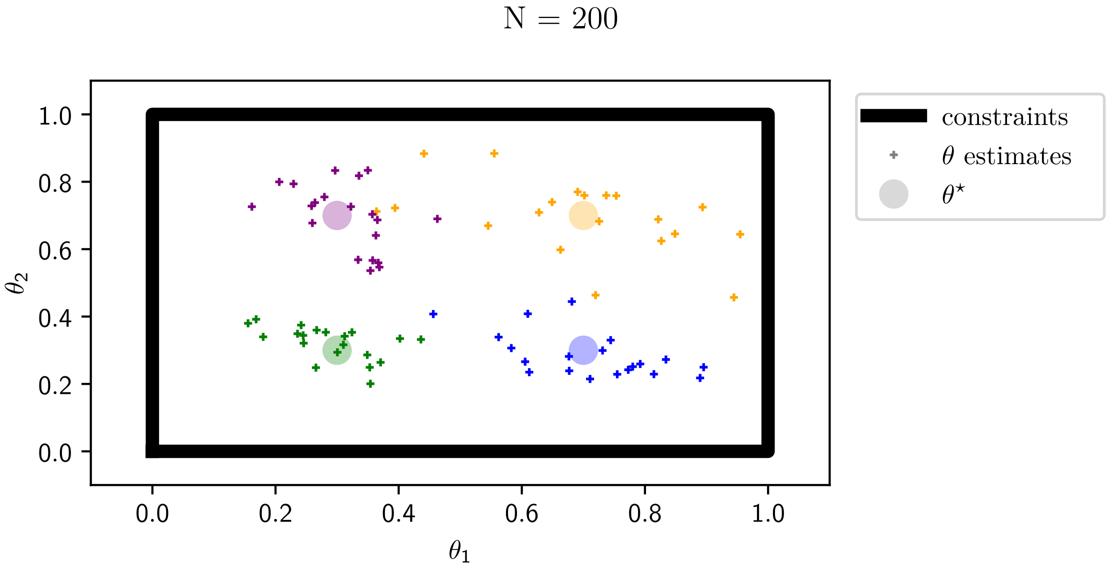


Performance on the Random Walk example



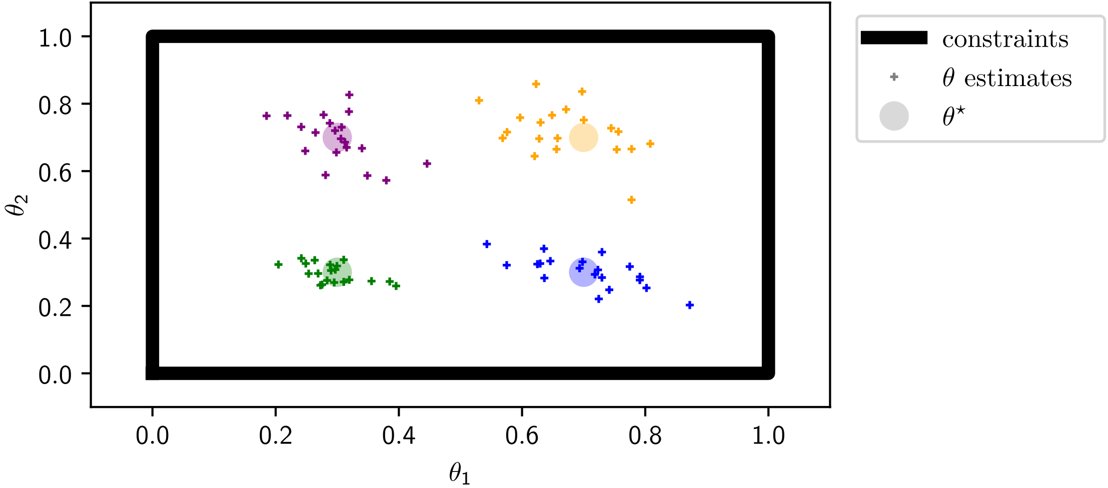

N = 100
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
θ_2
θ_1
constraints
θ estimates
$\theta^\star$

Performance on the Random Walk example

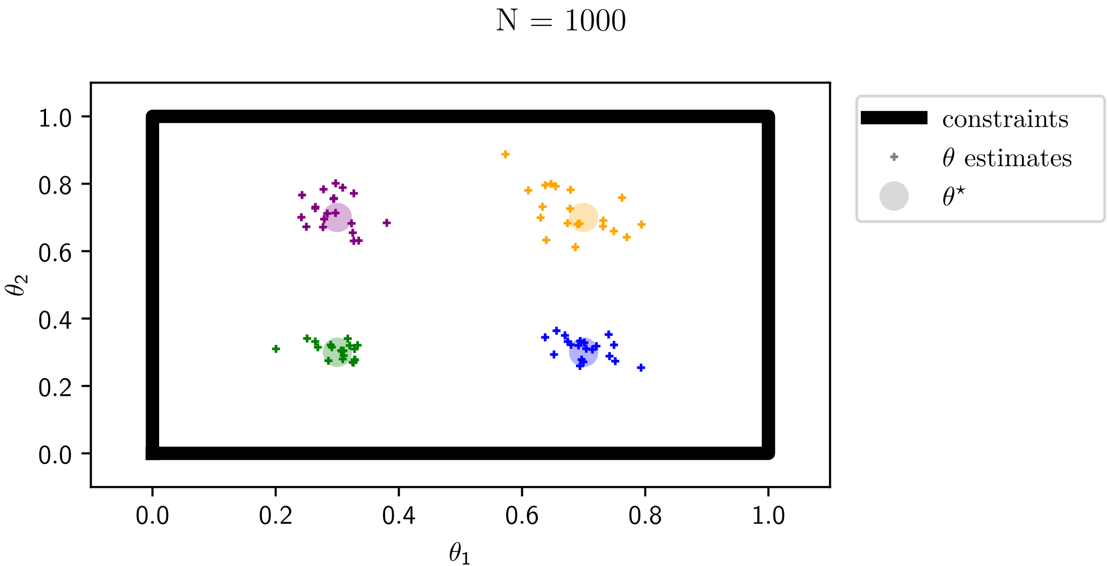


Performance on the Random Walk example

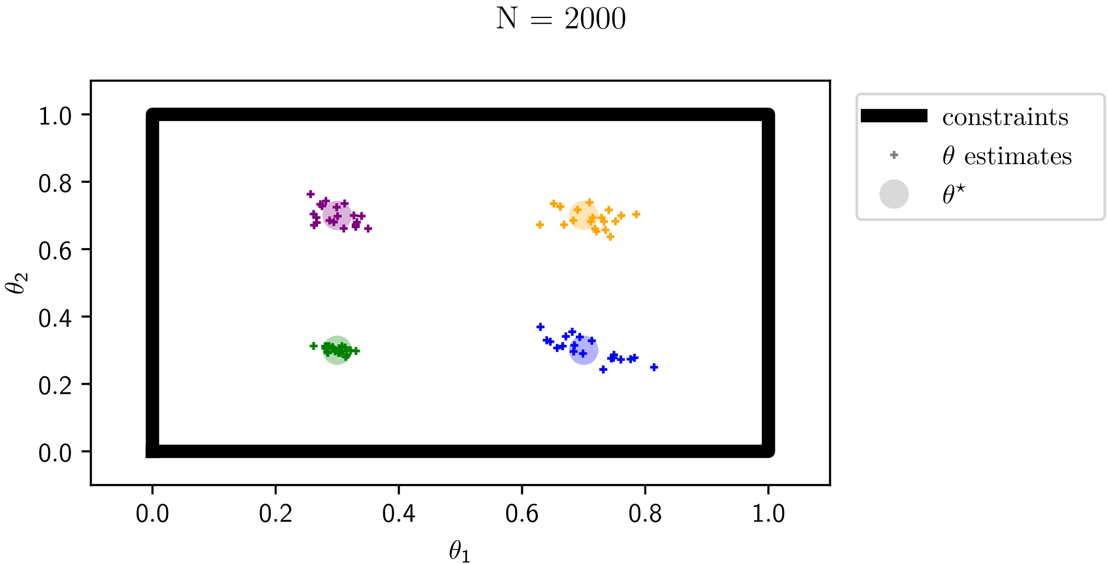
$\mathrm{N} = 500$



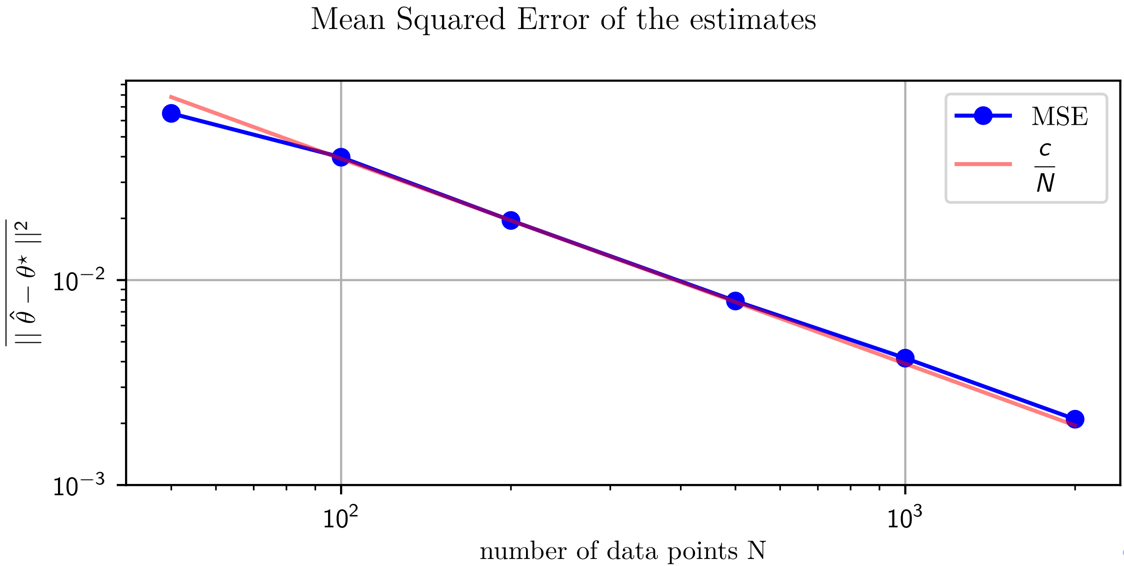
Performance on the Random Walk example



Performance on the Random Walk example



Performance on the Random Walk example



Conclusion

Conclusion

- The presented method has capabilities of joint estimation of parameters in the system dynamics and the noise covariances.
- Even with noisy data, it can recover the right parameters when the amount of data goes to infinite.
- We propose a change of paradigm: from "State and Parameter estimation" to "Identification of a State-Estimator".
- Thank you for your attention !