KU LEUVEN

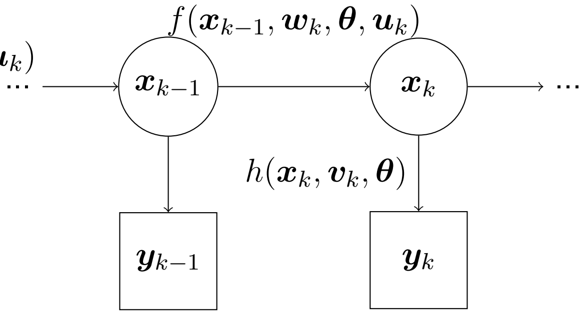
# Adaptive And Robust Experimental Design For Linear Dynamic Models Using The Kalman Filter

Arno Strouwen, Bart Nicolaï, Peter Goos

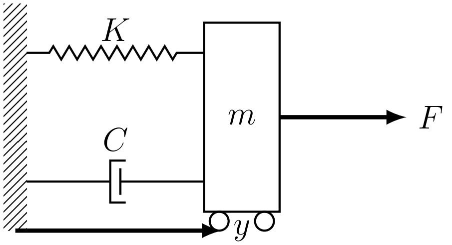
July 2023

# Dynamic systems

- Dynamic state $\boldsymbol{x}_k$ at time $t_k$
- Evolves in time $\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}, \boldsymbol{w}_k, \boldsymbol{\theta}, \boldsymbol{u}_k)$
  - Process noise $\boldsymbol{w}_k$
  - Unknown (static) parameters $\boldsymbol{\theta}$
  - Controllable input $\boldsymbol{u}_k$
- Measurements $\boldsymbol{y}_k = h(\boldsymbol{x}_k, \boldsymbol{v}_k, \boldsymbol{\theta})$
  - Measurement noise $\boldsymbol{v}_k$



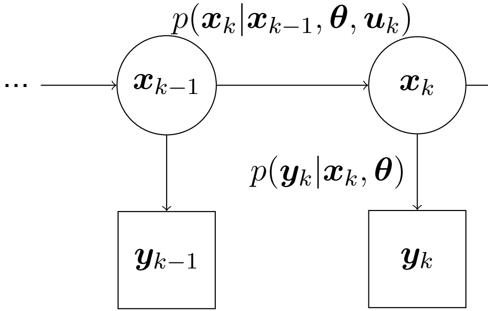
# Example dynamic system




- Mass spring damper system
- 2 states
  - Position $x_1$
  - Velocity $x_2$
- Change in position
  - $(x_1)_{k+1} = (x_1)_k + (x_2)_k \Delta t$
- Change in velocity by Newton's second law
  - $m\frac{(x_2)_{k+1} - (x_2)_k}{\Delta t} = -K(x_1)_k - C(x_2)_k + F +$ process noise
  - Pulling force F is controllable input
- Measurement$(y)$ = position$(x_1)$ + measurement noise
- K and C will be the unknown parameters

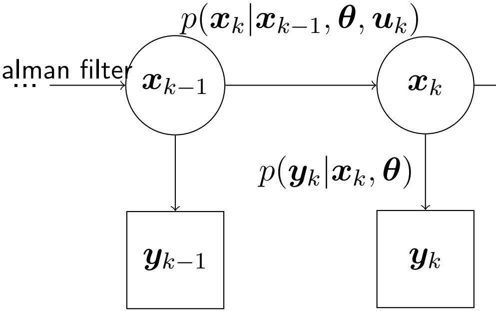
# Hidden Markov model

- Dynamic model $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}, \boldsymbol{u}_k)$
- Measurement model $p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta})$
- Markov Properties
  - Current state given previous state and input independent of anything that happened before $p(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1}, \boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:k}) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}, \boldsymbol{u}_k)$
  - Current measurement given previous state independent of anything that happened before $p(\boldsymbol{y}_k|\boldsymbol{x}_{1:k}, \boldsymbol{y}_{1:k-1}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta})$



# Sketch of Parameter estimation

- We want Likelihood of parameters $p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})$
  - Does not contain $\boldsymbol{x}$
  - Write in terms of Bayesian filtering equations for state estimation:
    State prediction distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})$
    State filtering distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})$
  - To estimate parameters, must estimate state
  - Optimal filter for linear dynamic systems is the Kalman filter



# Parameter estimation

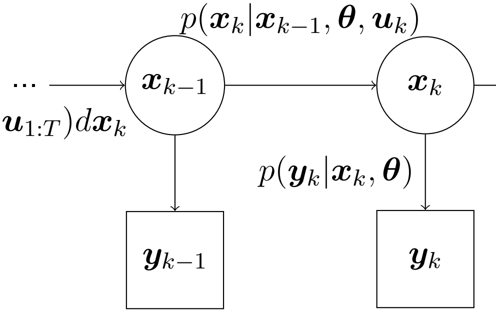
- Likelihood $p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})$
  - Write in terms of: $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}, \boldsymbol{u}_k)$ and $p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta})$

  $= p(\boldsymbol{y}_T|\boldsymbol{y}_{1:(T-1)}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})p(\boldsymbol{y}_{1:(T-1)}|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})$

  $= \prod_{k=1}^{T} p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})$
- $p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})$

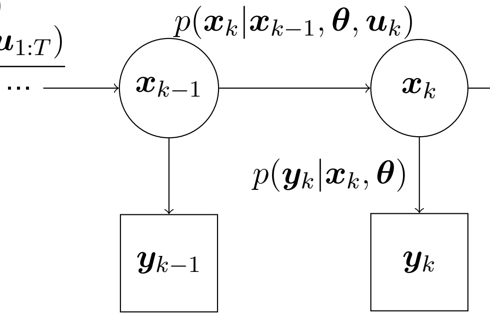
  $= \int p(\boldsymbol{y}_k, \boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})d\boldsymbol{x}_k$

  $= \int p(\boldsymbol{y}_k|\boldsymbol{x}_k, \cancel{\boldsymbol{y}_{1:k-1}}, \boldsymbol{\theta}, \cancel{\boldsymbol{u}_{1:T}})p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T})d\boldsymbol{x}_k$



# Bayesian filtering

- State prediction distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta},\boldsymbol{u}_{1:T})$
  $= \int p(\boldsymbol{x}_k,\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta},\boldsymbol{u}_{1:T})d\boldsymbol{x}_{k-1}$
  $= \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\cancel{\boldsymbol{y}_{1:k-1}},\boldsymbol{\theta},\boldsymbol{u}_{\substack{\cancel{1:T}\\ k}})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta},\boldsymbol{u}_{1:T})d\boldsymbol{x}_{k-1}$
- State filtering distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k},\boldsymbol{\theta},\boldsymbol{u}_{1:T})$
  $$= \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k,\cancel{\boldsymbol{y}_{1:k-1}},\boldsymbol{\theta},\cancel{\boldsymbol{u}_{1:T}})p(\boldsymbol{x}_k|\boldsymbol{y}_{1:(k-1)},\boldsymbol{\theta},\boldsymbol{u}_{1:T})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta},\boldsymbol{u}_{1:T})}$$
- Recurse to the state prior: $p(\boldsymbol{x}_0|\boldsymbol{\theta})$
  filtering distribution at k requires
  prediction distribution at k requires
  filtering distribution at k-1 requires
- Normalization factor $p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1},\boldsymbol{\theta},\boldsymbol{u}_{1:T})$
  Also what we needed for likelihood



# Linear dynamic systems

- $$\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}, \boldsymbol{u}_k, \boldsymbol{w}_k, \boldsymbol{\theta})$$
  $$\boldsymbol{y}_k = h(\boldsymbol{x}_k, \boldsymbol{v}_k, \boldsymbol{\theta})$$
- becomes
  $$\boldsymbol{x}_k = F(\boldsymbol{\theta})\boldsymbol{x}_{k-1} + B(\boldsymbol{\theta})\boldsymbol{u}_k + \boldsymbol{w}_k$$
  $$\boldsymbol{y}_k = H(\boldsymbol{\theta})\boldsymbol{x}_k + \boldsymbol{v}_k$$
  $$\boldsymbol{w}_k \sim \mathcal{N}(\mathbf{0}, Q(\boldsymbol{\theta})) \qquad \forall k$$
  $$\boldsymbol{v}_k \sim \mathcal{N}(\mathbf{0}, R(\boldsymbol{\theta})) \qquad \forall k$$
  $$\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{m}_0, P_0)$$
  $$\mathrm{Covar}(\boldsymbol{w}_k, \boldsymbol{v}_l) = \mathbf{0} \qquad \forall k, l$$
  $$\mathrm{Covar}(\boldsymbol{w}_k, \boldsymbol{w}_l) = \mathrm{Covar}(\boldsymbol{v}_k, \boldsymbol{v}_l) = \mathbf{0} \qquad \forall k, l \qquad k \neq l$$
  $$\mathrm{Covar}(\boldsymbol{x}_0, \boldsymbol{v}_k) = \mathrm{Covar}(\boldsymbol{x}_0, \boldsymbol{w}_k) = \mathbf{0} \qquad \forall k$$
- Gaussian distributions remain Gaussian under linear transformations

# Mass-Spring-Damper is linear

- $\boldsymbol{x}_k = \begin{bmatrix} 1 & \Delta t \\ -K\Delta t/m & 1 - c\Delta t/m \end{bmatrix} \boldsymbol{x}_{k-1} + \begin{bmatrix} 0 \\ \Delta t/m \end{bmatrix} \boldsymbol{u}_k + \boldsymbol{w}_k$
- $\boldsymbol{y}_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{x}_k + \boldsymbol{v}_k$
- $\boldsymbol{w}_k \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1000\Delta t/m & 0 \\ 0 & 1000\Delta t/m \end{bmatrix}\right)$
- $\boldsymbol{v}_k \sim \mathcal{N}(0, 100)$
- $\boldsymbol{P}_0 \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$

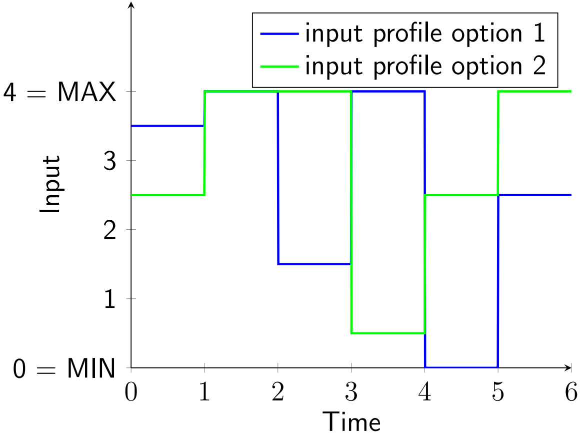
# Kalman filter

- Bayesian filter for linear systems
- $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}, \boldsymbol{\theta}, \boldsymbol{u}_{1:k}) \sim \mathcal{N}(\boldsymbol{m}_k, P_k)$
- prediction
  $$\boldsymbol{m}_k^- = F(\boldsymbol{\theta})\boldsymbol{m}_{k-1} + B(\boldsymbol{\theta})\boldsymbol{u}_k$$
  $$P_k^- = F(\boldsymbol{\theta})P_k F'(\boldsymbol{\theta}) + Q(\boldsymbol{\theta})$$
- update step
  $$\boldsymbol{v}_k = \boldsymbol{y}_k - H(\boldsymbol{\theta})\boldsymbol{m}_k^-$$
  $$S_k = H(\boldsymbol{\theta})P_k^- H'(\boldsymbol{\theta}) + R(\boldsymbol{\theta})$$
  $$K_k = P_k^- H'(\boldsymbol{\theta})S_k^{-1}$$
  $$\boldsymbol{m}_k = \boldsymbol{m}_k^- + K_k\boldsymbol{v}_k$$
  $$P_k = P_k^- - K_k S_k K'$$
- Likelihood $\prod_k p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1}, \boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = \prod_k \left(1/\sqrt{|2\pi S_k|} + \exp(-\frac{1}{2}\boldsymbol{v}_k' S_k^{-1}\boldsymbol{v}_k)\right)$


KU LEUVEN

# Optimal experimental design

- Optimize inputs $\boldsymbol{u}_{1:T}$ to estimate $\boldsymbol{\theta}$ as precisely as possible.



# Optimal experimental design

- Optimize inputs $\boldsymbol{u}_{1:T}$ to estimate $\boldsymbol{\theta}$ as precisely as possible.
- Expected Fisher information matrix

  $$\mathcal{I}(\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = -E_{\boldsymbol{y}_{1:T}|\boldsymbol{\theta},\boldsymbol{u}_{1:T}} \frac{\partial^2 \log p(\boldsymbol{y}_{1:T}|\boldsymbol{\theta},\boldsymbol{u}_{1:T})}{\partial \boldsymbol{\theta}^2}$$

  - Intuitive interpretation: sharp likelihood on average over all possible measurements
- The i,j'th entry of $\mathcal{I}_{i,j}(\boldsymbol{\theta}, \boldsymbol{u}_{1:T})$

  $$= \frac{\partial E(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}_i} \mathrm{Covar}(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})^{-1} \frac{\partial E(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} +$$

  $$\frac{1}{2} \mathrm{Tr} \left( \mathrm{Covar}(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})^{-1} \frac{\partial\, \mathrm{Covar}(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} \mathrm{Covar}(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})^{-1} \frac{\partial\, \mathrm{Covar}(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \right)$$
- D-optimality

# Recursion to calculate expected FIM

$$\boldsymbol{x}_k = F(\boldsymbol{\theta})\boldsymbol{x}_{k-1} + B(\boldsymbol{\theta})\boldsymbol{u}_k + \boldsymbol{w}_k$$
$$\boldsymbol{y}_k = H(\boldsymbol{\theta})\boldsymbol{x}_k + \boldsymbol{v}_k$$

- $E(\boldsymbol{y}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = H(\boldsymbol{\theta})E(\boldsymbol{x}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})$
  $E(\boldsymbol{x}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = F(\boldsymbol{\theta})E(\boldsymbol{x}_{k-1}|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) + B(\boldsymbol{\theta})\boldsymbol{u}_k$
  $E(\boldsymbol{x}_0|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = \boldsymbol{m}_0$
  $\mathrm{Var}(\boldsymbol{y}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = H(\boldsymbol{\theta})\,\mathrm{Var}(\boldsymbol{x}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})H(\boldsymbol{\theta})' + R(\boldsymbol{\theta})$
  $\mathrm{Var}(\boldsymbol{x}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = F(\boldsymbol{\theta})\,\mathrm{Var}(\boldsymbol{x}_{k-1}|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})F(\boldsymbol{\theta})' + Q(\boldsymbol{\theta})$
  $\mathrm{Var}(\boldsymbol{x}_0|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = P_0$
  $\mathrm{Covar}(\boldsymbol{y}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}; \boldsymbol{y}_l|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = H(\boldsymbol{\theta})F^{k-l}\,\mathrm{Var}(\boldsymbol{x}_l|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})H(\boldsymbol{\theta})', \quad \forall k > l$
  $\mathrm{Covar}(\boldsymbol{y}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}; \boldsymbol{y}_l|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}) = \mathrm{Covar}(\boldsymbol{y}_l|\boldsymbol{\theta}, \boldsymbol{u}_{1:T}; \boldsymbol{y}_k|\boldsymbol{\theta}, \boldsymbol{u}_{1:T})' \quad \forall k < l$
  - Related to repeatedly using the prediction step of the Kalman filter
- Can also derive recursion for parameter sensitivities
  $\frac{\partial E(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}_i}, \frac{\partial\,\mathrm{Covar}(\boldsymbol{y}_{1:T}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}$
  - Automatic differentiation more convenient

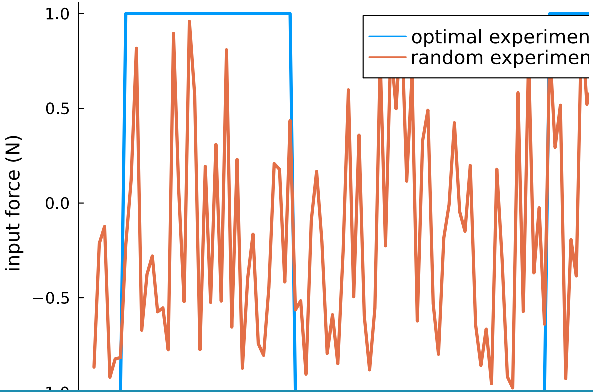
# Robust & adaptive experiments

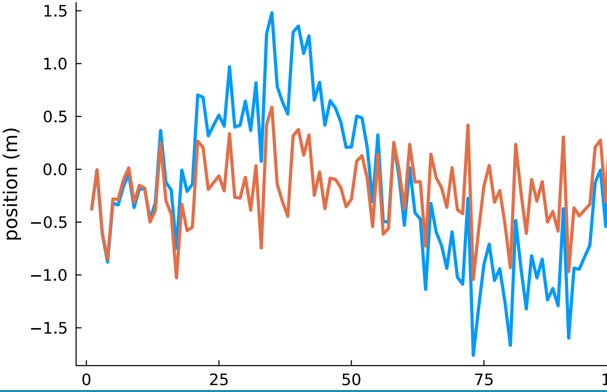
- Expected FIM depends on the unknown parameters
- Average out D-criterion over prior distribution
  $\arg\max_{\boldsymbol{u}_{1:T}} \int |\mathcal{I}(\boldsymbol{\theta}, \boldsymbol{u}_{1:T})| \, p(\boldsymbol{\theta}) d\boldsymbol{\theta}$
- Adapt experimental plan after every measurement
  $\arg\max_{\boldsymbol{u}_{(k+1):T}} \int |\mathcal{I}(\boldsymbol{\theta}, \boldsymbol{u}_{1:T})| \, p(\boldsymbol{\theta}|\boldsymbol{y}_{1:k}) d\boldsymbol{\theta}$

# Adaptive horizon

- But correlations between $\boldsymbol{y}$ not feasible online.
- Observed FIM $\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{u}_{1:k}) = \frac{\partial \log p(\boldsymbol{y}_{1:k}|\boldsymbol{\theta}, \boldsymbol{u}_{1:k})}{\partial \boldsymbol{\theta}} \frac{\partial \log p(\boldsymbol{y}_{1:k}|\boldsymbol{\theta}, \boldsymbol{u}_{1:k})}{\partial \boldsymbol{\theta}}'$
- Prediction far into the future too difficult
  $\arg\max_{\boldsymbol{u}_{(k+1):(k+e)}} \int \left|\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{u}_{1:k}) + \mathcal{I}(\boldsymbol{\theta}, \boldsymbol{u}_{(k+1):(k+e)})\right| p(\boldsymbol{\theta}|\boldsymbol{y}_{1:k}) d\boldsymbol{\theta}$
- Current limitation: Monte Carlo integration has to be performed at the same $\boldsymbol{\theta}_i$ every timestep.

# Input profile and corresponding output

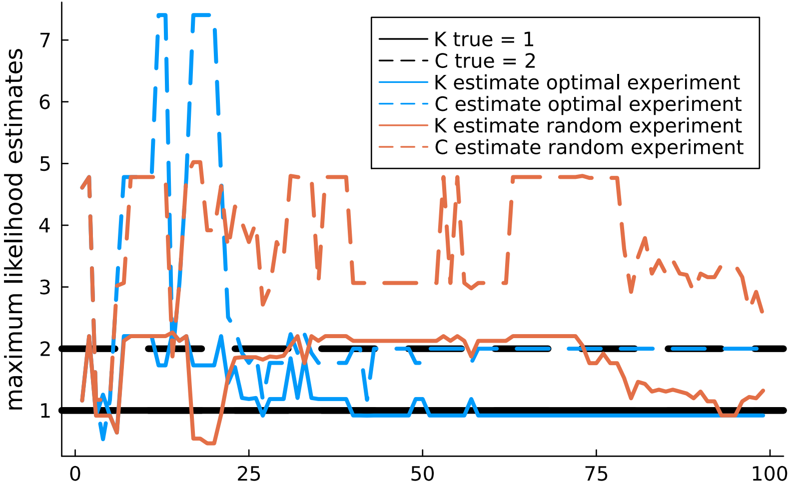


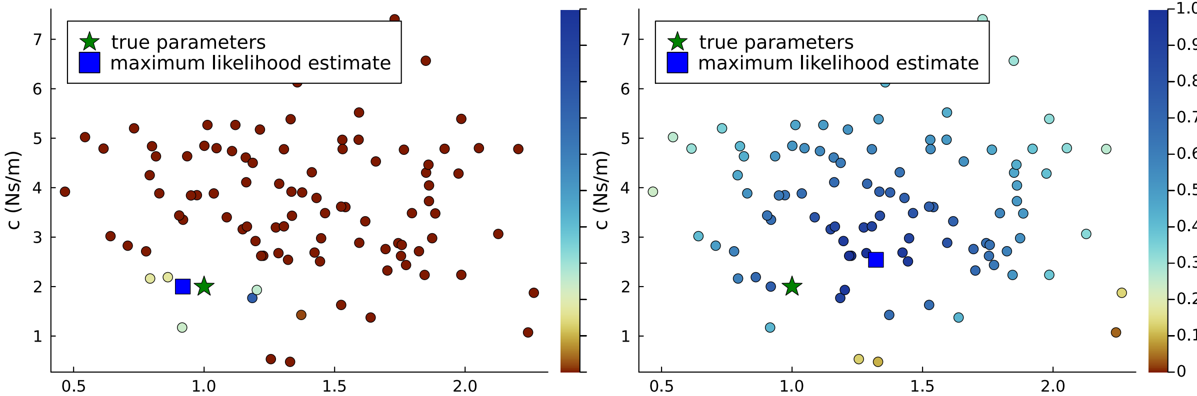



KU LEUVEN

# Online maximum likelihood estimate



# Likelihood at end of experiment






KU LEUVEN