Non-Parametric Bayesian Methods for Linear System Identification

Giulia Prando

PhD Advisor: Prof. Alessandro Chiuso
PhD Co-Advisor: Prof. Gianluigi Pillonetto

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Department of Information Engineering - University of Padova

24 March 2017

System

System

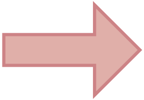
Inputs

u

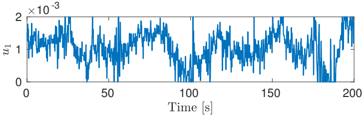
System

Inputs

u



slip on a tire

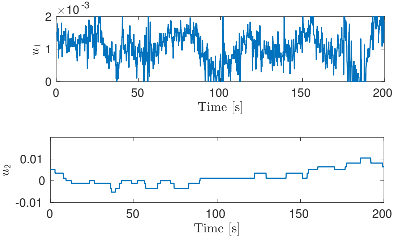


System

Inputs

u

slip on a tire
steering angle



System vs Model

System

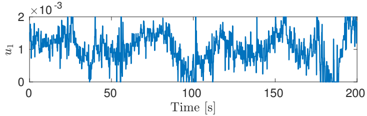
Inputs

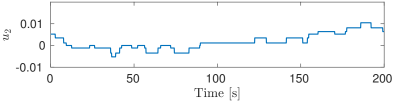
u

slip on a tire
steering angle

Outputs

y





System

Inputs

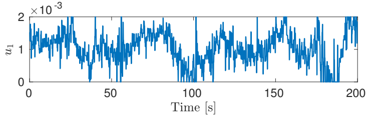
u

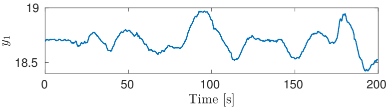
slip on a tire
steering angle

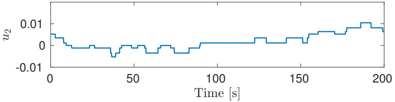
Outputs

y

longitudinal velocity







System

Inputs

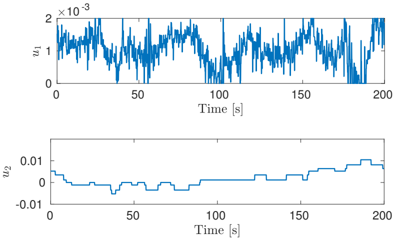
u

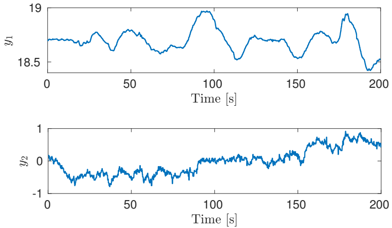
slip on a tire
steering angle

Outputs

y

longitudinal velocity
lateral acceleration





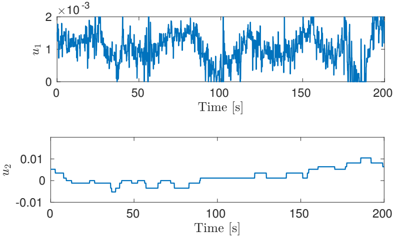
Model

$$\sum_{k=1}^{\infty} h(k)u(t-k)$$

Inputs

u

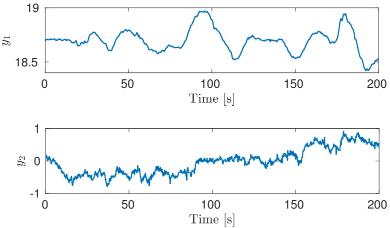
slip on a tire
steering angle

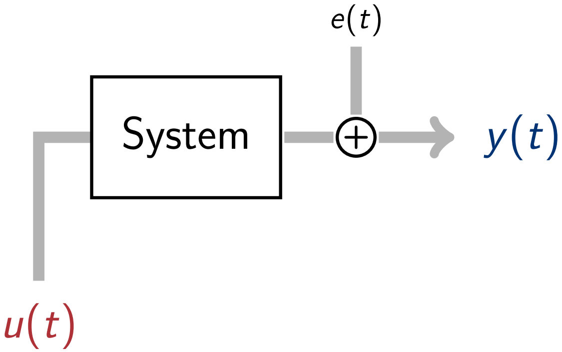


Outputs

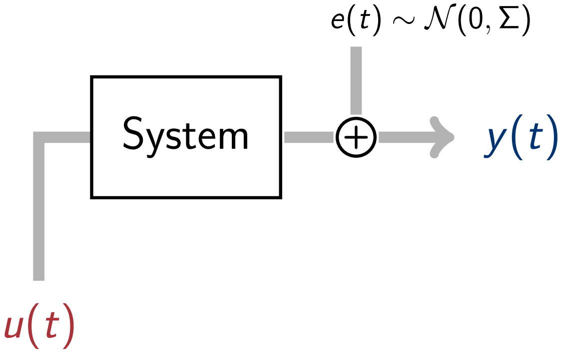
y

longitudinal velocity
lateral acceleration

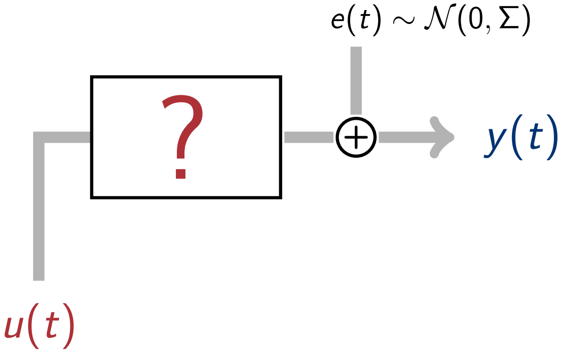




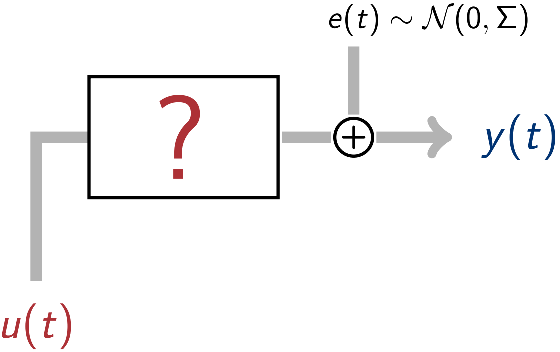
$t = 1, ..., N$



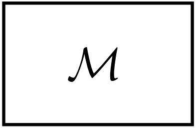
$t = 1, \ldots, N$



$t = 1, ..., N$

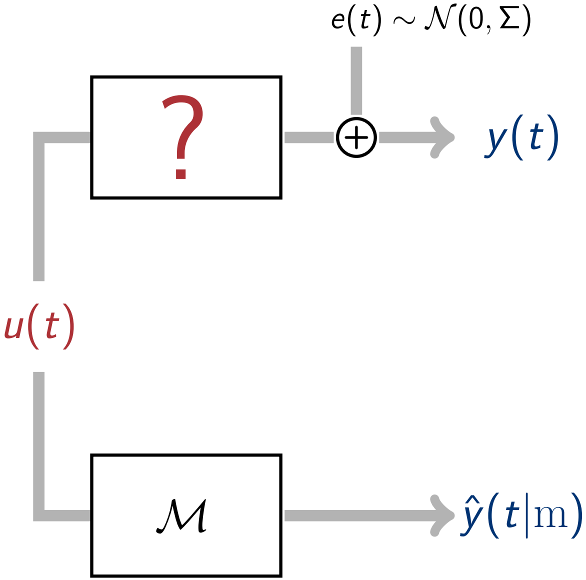


$t = 1, ..., N$

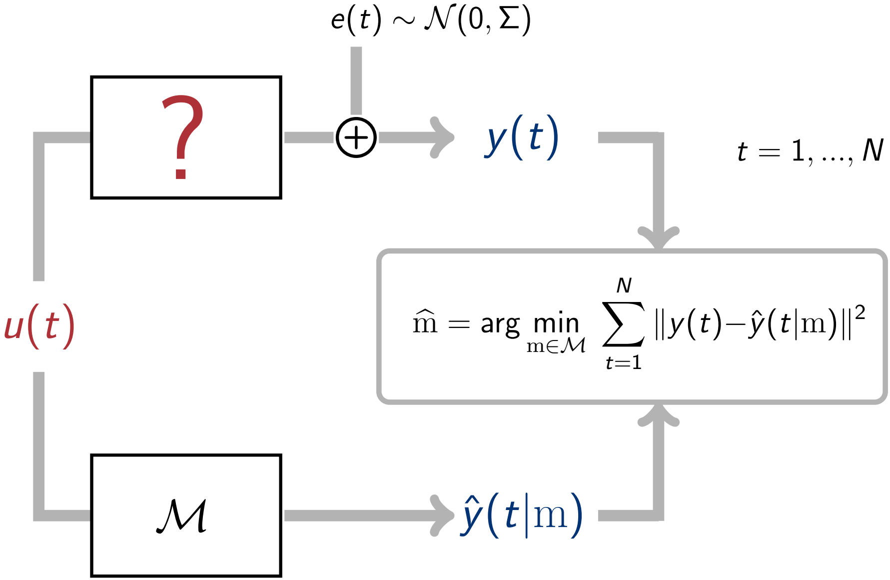


System Identification Approach

Ljung 1999



$t = 1, \ldots, N$



$e(t) \sim \mathcal{N}(0, \Sigma)$
?
$y(t)$
$t = 1, \dots, N$
$u(t)$
$\widehat{\mathrm{m}} = \arg \min_{\mathrm{m} \in \mathcal{M}} \sum_{t=1}^{N} \|y(t) - \hat{y}(t|\mathrm{m})\|^2$
$\mathcal{M}$
$\hat{y}(t|\mathrm{m})$

Two families of methods

Parametric

Non-Parametric

Two families of methods

Parametric	Non-Parametric
• PEM • Subspace	

Two families of methods

Parametric

- PEM
- Subspace

- $\mathcal{M}(\theta)$, $\theta \in \mathcal{R}^d$

  E.g. $$\mathcal{M}(\theta) = \frac{\sum_{i=1}^{n} \theta_i q^{-i}}{1 + \sum_{j=n+1}^{d} \theta_j q^{-j}}$$

- Estimate θ

Non-Parametric

Two families of methods

Parametric

- PEM
- Subspace

- $\mathcal{M}(\theta)$, $\theta \in \mathcal{R}^d$

E.g. $$\mathcal{M}(\theta) = \frac{\sum_{i=1}^{n} \theta_i q^{-i}}{1 + \sum_{j=n+1}^{d} \theta_j q^{-j}}$$

- Estimate θ

- How many parameters?
- Complexity Selection Issue

Non-Parametric

Two families of methods

Parametric

- PEM
- Subspace

- $\mathcal{M}(\theta)$, $\theta \in \mathcal{R}^d$

E.g. $$\mathcal{M}(\theta) = \frac{\sum_{i=1}^{n} \theta_i q^{-i}}{1 + \sum_{j=n+1}^{d} \theta_j q^{-j}}$$

- Estimate θ

- How many parameters?
- Complexity Selection Issue

Non-Parametric

This talk

Two families of methods

Parametric

- PEM
- Subspace

- $\mathcal{M}(\theta)$, $\theta \in \mathcal{R}^d$

E.g. $$\mathcal{M}(\theta) = \frac{\sum_{i=1}^{n} \theta_i q^{-i}}{1 + \sum_{j=n+1}^{d} \theta_j q^{-j}}$$

- Estimate θ

- How many parameters?
- Complexity Selection Issue

Non-Parametric

This talk

Directly estimate
impulse response h

Parametric

- PEM
- Subspace

- $\mathcal{M}(\theta)$, $\theta \in \mathcal{R}^d$

E.g. $$\mathcal{M}(\theta) = \frac{\sum_{i=1}^{n} \theta_i q^{-i}}{1 + \sum_{j=n+1}^{d} \theta_j q^{-j}}$$

- Estimate θ

- How many parameters?
- Complexity Selection Issue

Non-Parametric

This talk

Directly estimate
impulse response h

- No complexity selection needed
- Exploit the Bayesian setting to include a-priori information

Non-Parametric Bayesian Methods

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Non-Parametric Bayesian Methods

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

- Accounts for desired properties of h
- Depends on hyper-parameters η

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

- Accounts for desired properties of h
- Depends on hyper-parameters η

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

- Inference step using the observed data $\mathbf{y}$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

- Accounts for desired properties of h
- Depends on hyper-parameters η

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

- Inference step using the observed data $\mathbf{y}$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

- Minimum variance estimator
- Analytically intractable

$\Downarrow$

Approximations required

My Research Activity

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \ p_\eta(h|\mathbf{y})dh$$

Post-Processing

$$\hat{h} \longrightarrow h(\theta), \ \theta \in \mathbb{R}^d$$

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Prior accounting for complexity

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Post-Processing

$$\hat{h} \longrightarrow h(\theta), \;\; \theta \in \mathbb{R}^d$$

Prando, Chiuso, and Pillonetto 2017
Prando, Pillonetto, and Chiuso 2015
Prando, Chiuso, and Pillonetto 2014

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Assessment of estimator uncertainty

Prando et al. 2016

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Post-Processing

$$\hat{h} \longrightarrow h(\theta), \;\; \theta \in \mathbb{R}^d$$

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Prando et al. 2016

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Comparison of 2 approximations:

- Empirical Bayes
- Full Bayes (using sampling methods)

Post-Processing

$$\hat{h} \longrightarrow h(\theta), \;\; \theta \in \mathbb{R}^d$$

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Post-Processing

$$\hat{h} \longrightarrow h(\theta), \;\; \theta \in \mathbb{R}^d$$

Online Re-formulation:

Update the estimate
as soon as
new data arrive

Prando, Romeres, and Chiuso 2016
Romeres et al. 2016

My Research Activity

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Prando and Chiuso 2015

Post-Processing

$$\hat{h} \longrightarrow h(\theta), \;\; \theta \in \mathbb{R}^d$$

Model reduction

Outline

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h \; p_\eta(h|\mathbf{y})dh$$

Post-Processing

$$\hat{h} \longrightarrow h(\theta)$$

Outline

Prior

$$h \sim p_\eta(h) = \mathcal{N}(0, K_\eta)$$

Prior accounting for complexity

Posterior

$$p_\eta(h|\mathbf{y}) = \frac{p(\mathbf{y}|h)p_\eta(h)}{p_\eta(\mathbf{y})}$$

Point Estimator

$$\hat{h} = \mathbb{E}_\eta[h|\mathbf{y}] = \int h\ p_\eta(h|\mathbf{y})dh$$

Post-Processing

$$\hat{h} \longrightarrow h(\theta)$$

Smoothness

Smoothness

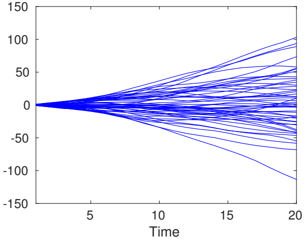
- Smoothing-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(s,t)$$

Smoothness

- Smoothing-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(s,t)$$

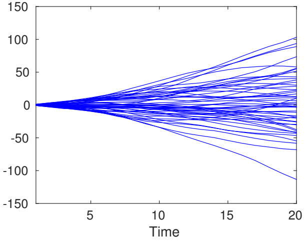


Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(s,t))$

Smoothness

- Smoothing-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(s,t)$$



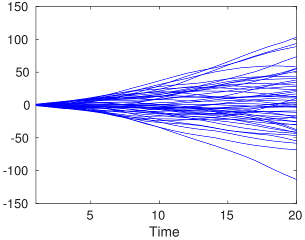
Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(s,t))$

Smoothness and Stability

Smoothness

- Smoothing-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(s,t)$$



Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(s,t))$

Smoothness and Stability

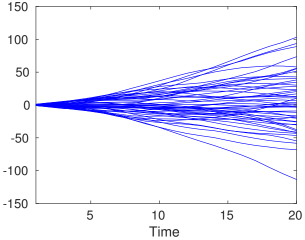
- Stable-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(\beta^s, \beta^t)$$

Smoothness

- Smoothing-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(s,t)$$

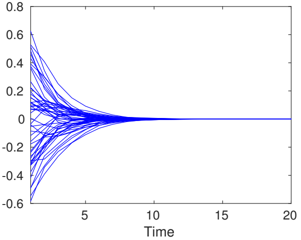


Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(s,t))$

Smoothness and Stability

- Stable-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(\beta^s, \beta^t)$$

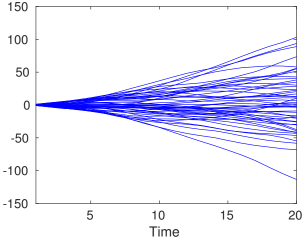


Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(\beta^s, \beta^t))$

Smoothness

- Smoothing-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(s,t)$$

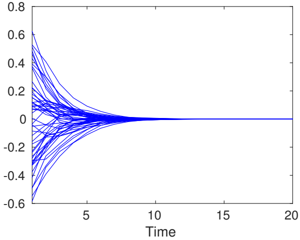


Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(s,t))$

Smoothness and Stability

- Stable-spline kernel

$$K_\eta(s,t) = K_\eta^{SS}(\beta^s, \beta^t)$$



Realizations from $p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SS}(\beta^s, \beta^t))$

Remark: Stable spline kernels don't account for model complexity

Aim:

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Recall:

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Recall:

$$\mathcal{H}(h) = \begin{bmatrix} h(1) & h(2) & h(3) & \cdots & h(c) \\ h(2) & h(3) & h(4) & \cdots & h(c+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(r) & h(r+1) & \cdots & \cdots & h(r+c-1) \end{bmatrix}$$

$rank(\mathcal{H}(h))$ = Mc Millan degree of the system with impulse response h

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Complexity Constraint:

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Complexity Constraint:

$$\mathbb{E}\left[\sum_i \frac{s_i\left\{\mathcal{H}(h)\mathcal{H}(h)^\top\right\}}{\alpha_i}\right] \leq c_C$$

Aim:

Build Max-Entropy prior $p_\eta(h)$ satisfying constraints on:

- Stability
- Complexity

Stability Constraint:

$$\mathbb{E}\left[h^\top K_\rho{}^{-1} h\right] \leq c_S$$

Satisfied if K_ρ is the **stable-spline kernel** [Pillonetto and De Nicolao 2010]

Complexity Constraint:

$$\mathbb{E}\left[h^\top P^\top (Q_\alpha \otimes I) P h\right] \leq c_C$$

With a properly designed matrix Q_α

Problem:

Problem:

$$p_\eta(h) = \arg\max_p \; H[p(h)]$$

$$s.t. \quad \mathbb{E}\left[h^\top K_\rho^{-1} h\right] \leq c_S \qquad \textbf{Stability}$$

$$\mathbb{E}\left[h^\top P^\top (Q_\alpha \otimes I) P h\right] \leq c_C \qquad \textbf{Low complexity}$$

Problem:

$$p_\eta(h) = \arg\max_p \; H[p(h)]$$

$$s.t. \quad \mathbb{E}\left[h^\top K_\rho^{-1} h\right] \leq c_S \qquad \textbf{Stability}$$

$$\mathbb{E}\left[h^\top P^\top (Q_\alpha \otimes I) P h\right] \leq c_C \qquad \textbf{Low complexity}$$

Solution:

$$p_\eta(h) \propto \exp\left(-\frac{\lambda_S}{2} h^\top K_\rho^{-1} h - \frac{\lambda_H}{2} h^\top P^\top (Q_\alpha \otimes I) P h\right)$$

Problem:

$$p_\eta(h) = \arg\max_p \; H[p(h)]$$

$$s.t. \quad \mathbb{E}\left[h^\top K_\rho^{-1} h\right] \le c_S \qquad \textbf{Stability}$$

$$\mathbb{E}\left[h^\top P^\top (Q_\alpha \otimes I) P h\right] \le c_C \qquad \textbf{Low complexity}$$

Solution:

$$p_\eta(h) \propto \exp\left(-\frac{\lambda_S}{2} h^\top K_\rho^{-1} h - \frac{\lambda_H}{2} h^\top P^\top (Q_\alpha \otimes I) P h\right)$$

Gaussian Prior:

$$p_\eta(h) = \mathcal{N}\left(0, \; \left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right]^{-1}\right)$$

Problem:

$$p_\eta(h) = \arg\max_p \; H[p(h)]$$

$$s.t. \quad \mathbb{E}\left[h^\top K_\rho^{-1} h\right] \leq c_S \qquad \textbf{Stability}$$

$$\mathbb{E}\left[h^\top P^\top (Q_\alpha \otimes I) P h\right] \leq c_C \qquad \textbf{Low complexity}$$

Solution:

$$p_\eta(h) \propto \exp\left(-\frac{\lambda_S}{2} h^\top K_\rho^{-1} h - \frac{\lambda_H}{2} h^\top P^\top (Q_\alpha \otimes I) P h\right)$$

Gaussian Prior:

$$p_\eta(h) = \mathcal{N}\left(0, \; \underbrace{\left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right]}_{\text{Stable-Hankel Kernel } K_\eta^{SH}}{}^{-1}\right)$$

Problem:

$$p_\eta(h) = \arg\max_{p} \; H[p(h)]$$

$$s.t. \quad \mathbb{E}\left[h^\top K_\rho^{-1} h\right] \leq c_S \qquad \textbf{Stability}$$

$$\mathbb{E}\left[h^\top P^\top (Q_\alpha \otimes I) P h\right] \leq c_C \qquad \textbf{Low complexity}$$

Solution:

$$p_\eta(h) \propto \exp\left(-\frac{\lambda_S}{2} h^\top K_\rho^{-1} h - \frac{\lambda_H}{2} h^\top P^\top (Q_\alpha \otimes I) P h\right)$$

Gaussian Prior:

$$p_\eta(h) = \mathcal{N}\left(0, \; \underbrace{\left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right]^{-1}}_{\text{Stable-Hankel Kernel} \;\; K_\eta^{SH}}\right)$$

Hyper-parameters: $\eta = \{\lambda_S, \; \lambda_H, \; \rho, \; \alpha\}$

Gaussian Prior

$$p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SH})$$

Gaussian Prior

$$p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SH})$$

Gaussian Prior

$$p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SH})$$

- Adopt Marginal Likelihood maximization to find the hyper-parameters

$$\hat{\eta} = \arg\max_\eta p_\eta(\mathbf{y})$$

Gaussian Prior

$$p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SH})$$

↓

- Adopt Marginal Likelihood maximization to find the hyper-parameters

$$\hat{\eta} = \arg\max_{\eta} p_\eta(\mathbf{y})$$

- Get $\hat{h}$ in closed-form

$$\hat{h} = \arg\min_{h} \sum_{t=1}^{N} \| y(t) - \hat{y}(t|h) \|_{\Sigma}^2 + h^\top \left(K_{\hat{\eta}}^{SH}\right)^{-1} h$$

Gaussian Prior

$$p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SH})$$

- Adopt Marginal Likelihood maximization to find the hyper-parameters

$$\hat{\eta} = \arg\max_{\eta} p_\eta(\mathbf{y})$$

- Get $\hat{h}$ in closed-form

$$\hat{h} = \arg\min_{h} \sum_{t=1}^{N} \|y(t) - \varphi^\top(t)h\|_\Sigma^2 + h^\top \left(K_{\hat{\eta}}^{SH}\right)^{-1} h$$

$$p_\eta(h) \sim \mathcal{N}(0, K_\eta^{SH})$$

- Adopt Marginal Likelihood maximization to find the hyper-parameters

$$\hat{\eta} = \arg\max_\eta p_\eta(\mathbf{y})$$

- Get $\hat{h}$ in closed-form

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \varphi^\top(t) h\|_\Sigma^2 + h^\top \left(K_{\hat{\eta}}^{SH}\right)^{-1} h$$

$$= \left[\sum_{t=1}^{N} \varphi(t)\Sigma^{-1}\varphi^\top(t) + \left(K_{\hat{\eta}}^{SH}\right)^{-1}\right]^{-1} \sum_{t=1}^{N} \varphi(t)\Sigma^{-1}y(t)$$

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_\Sigma^2 + h^\top \left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right] h$$

$$\hat{h} = \arg\min_{h} \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_{\Sigma}^{2} + h^{\top} \left[\lambda_S K_{\rho}^{-1} + \lambda_H P^{\top} (Q_{\alpha} \otimes I) P \right] h$$

- Assume $\lambda_S = 0$

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_{\Sigma}^2 + h^\top \left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right] h$$

- Assume $\lambda_S = 0 \implies$ No stability constraint

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_\Sigma^2 + h^\top \left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right] h$$

- Assume $\lambda_S = 0 \implies$ No stability constraint
- Recall

$$h^\top P(Q_\alpha \otimes I)Ph = \sum_{i=1}^{r} \frac{s_i\{\mathcal{H}(h)\mathcal{H}(h)^\top\}}{\alpha_i}$$

$$\hat{h} = \arg\min_{h} \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_{\Sigma}^{2} + h^{\top}\left[\lambda_S K_{\rho}^{-1} + \lambda_H P^{\top}(Q_{\alpha} \otimes I)P\right] h$$

- Assume $\lambda_S = 0 \quad \Longrightarrow$ No stability constraint
- Recall

$$h^{\top} P(Q_{\alpha} \otimes I)Ph = \sum_{i=1}^{r} \frac{s_i\{\mathcal{H}(h)\mathcal{H}(h)^{\top}\}}{\alpha_i}$$

- Assume $\alpha_1 = \cdots = \alpha_r = \alpha,$

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_\Sigma^2 + h^\top \left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right] h$$

- Assume $\lambda_S = 0 \quad \Longrightarrow \quad$ No stability constraint
- Recall

$$h^\top P(Q_\alpha \otimes I)Ph = \sum_{i=1}^{r} \frac{s_i\{\mathcal{H}(h)\mathcal{H}(h)^\top\}}{\alpha_i}$$

- Assume $\alpha_1 = \cdots = \alpha_r = \alpha,$ then

$$\sum_{i=1}^{r} \frac{s_i\{\mathcal{H}(h)\mathcal{H}(h)^\top\}}{\alpha_i} = \frac{1}{\alpha} \|\mathcal{H}(h)\mathcal{H}(h)^\top\|_*$$

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_\Sigma^2 + h^\top \left[\lambda_S K_\rho^{-1} + \lambda_H P^\top (Q_\alpha \otimes I) P\right] h$$

- Assume $\lambda_S = 0 \quad \Longrightarrow$ No stability constraint
- Recall

$$h^\top P(Q_\alpha \otimes I)Ph = \sum_{i=1}^{r} \frac{s_i\{\mathcal{H}(h)\mathcal{H}(h)^\top\}}{\alpha_i}$$

- Assume $\alpha_1 = \cdots = \alpha_r = \alpha,$ then

$$\sum_{i=1}^{r} \frac{s_i\{\mathcal{H}(h)\mathcal{H}(h)^\top\}}{\alpha_i} = \frac{1}{\alpha}\|\mathcal{H}(h)\mathcal{H}(h)^\top\|_*$$

Nuclear Norm Regularization:

$$\hat{h} = \arg\min_h \sum_{t=1}^{N} \|y(t) - \hat{y}(t|h)\|_\Sigma^2 + \frac{\lambda_H}{\alpha}\|\mathcal{H}(h)\mathcal{H}(h)^\top\|_*$$

Numerical Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Numerical Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Impulse Response Fit:

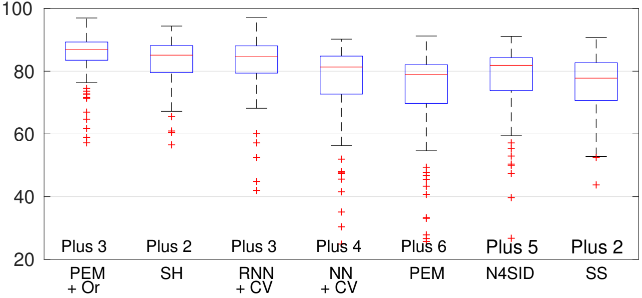
$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\|h_{ij} - \hat{h}_{ij}\|}{\|h_{ij}\|}\right)$$

Numerical Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$



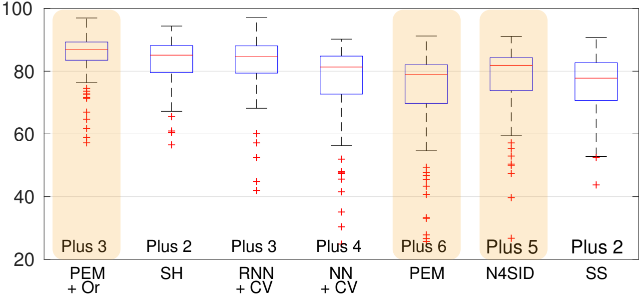
Median of $\mathcal{F}(\hat{h})$

Numerical Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Impulse Response Fit:

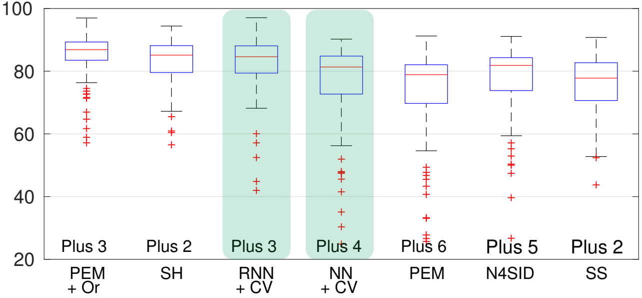
$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$



	Median of $\mathcal{F}(\hat{h})$
PEM+Or	86.7
PEM	78.9
N4SID	81.9

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Impulse Response Fit:
$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$

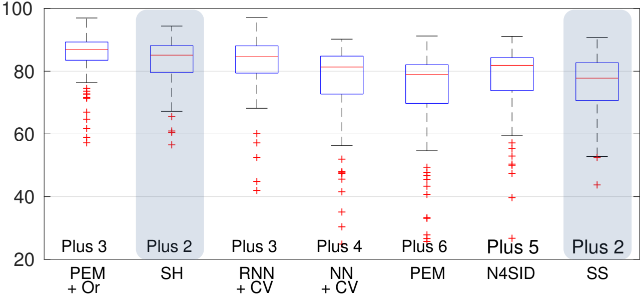


	Median of $\mathcal{F}(\hat{h})$
PEM+Or	86.7
RNN+CV	84.6
NN+CV	81.3
PEM	78.9
N4SID	81.9

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$



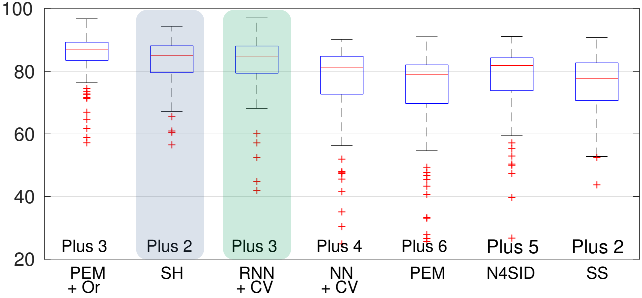
	Median of $\mathcal{F}(\hat{h})$
PEM+Or	86.7
SH	85.1
RNN+CV	84.6
NN+CV	81.3
PEM	78.9
N4SID	81.9
SS	77.8

Numerical Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\|h_{ij} - \hat{h}_{ij}\|}{\|h_{ij}\|} \right)$$

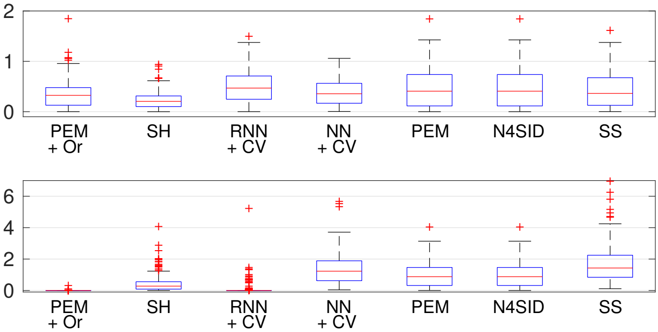


	Median of $\mathcal{F}(\hat{h})$
PEM+Or	86.7
SH	85.1
RNN+CV	84.6
NN+CV	81.3
PEM	78.9
N4SID	81.9
SS	77.8

Experimental Results

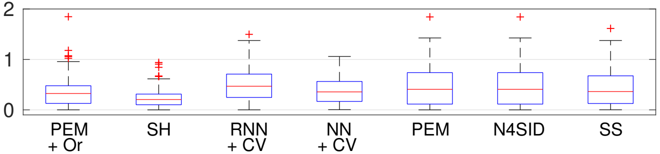
Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

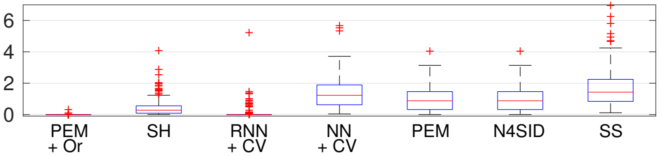
Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$



Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

$$\Delta_{signal}(\hat{h}) \quad := \quad \sum_{i=1}^{n} |s_i(h) \; - \; s_i(\hat{h})|$$

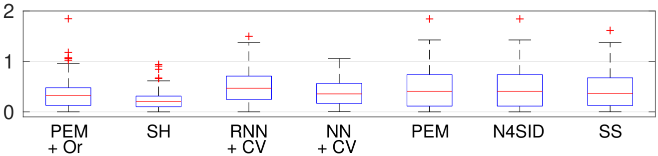


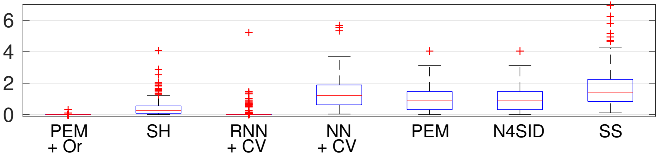


Experimental Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

$$\Delta_{signal}(\hat{h}) := \sum_{i=1}^{n} |s_i(h) - s_i(\hat{h})|$$



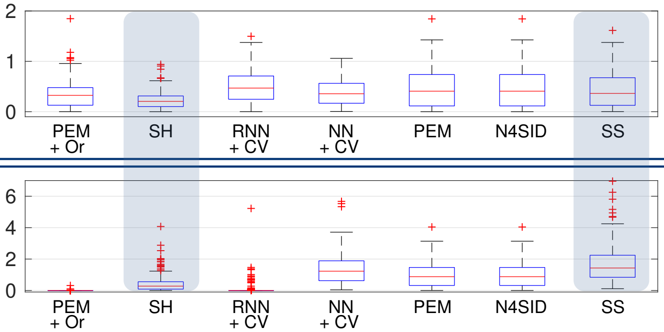


$$\Delta_{noise}(\hat{h}) := \sum_{i=n+1}^{r} |s_i(h) - s_i(\hat{h})|$$

Experimental Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

$$\Delta_{signal}(\hat{h}) := \sum_{i=1}^{n} |s_i(h) - s_i(\hat{h})|$$

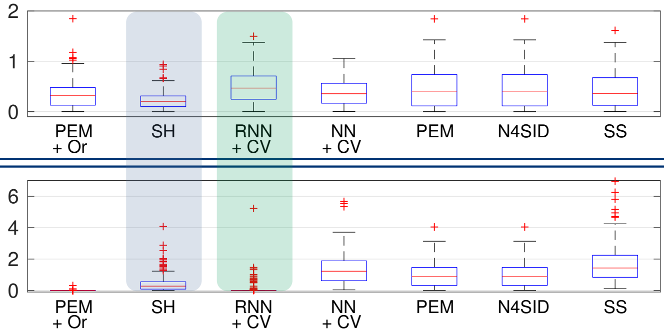


$$\Delta_{noise}(\hat{h}) := \sum_{i=n+1}^{r} |s_i(h) - s_i(\hat{h})|$$

Experimental Results

Setup: Random MIMO systems, 5 inputs/5 outputs, 200 MC runs, $N = 500$

$$\Delta_{signal}(\hat{h}) := \sum_{i=1}^{n} |s_i(h) - s_i(\hat{h})|$$



$$\Delta_{noise}(\hat{h}) := \sum_{i=n+1}^{r} |s_i(h) - s_i(\hat{h})|$$

Numerical Results

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

Numerical Results

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

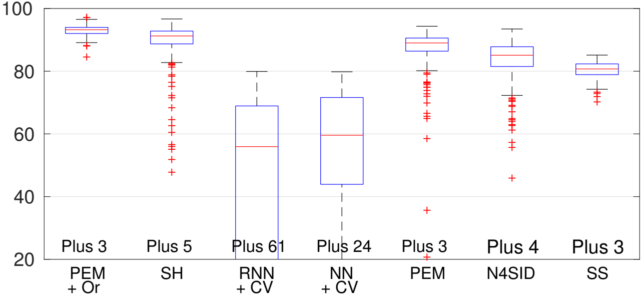
Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$



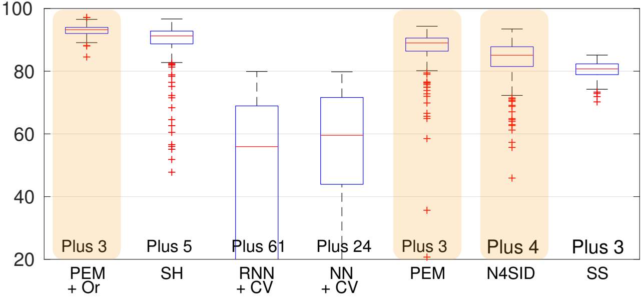
Median of $\mathcal{F}(\hat{h})$

Numerical Results

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\|h_{ij} - \hat{h}_{ij}\|}{\|h_{ij}\|} \right)$$

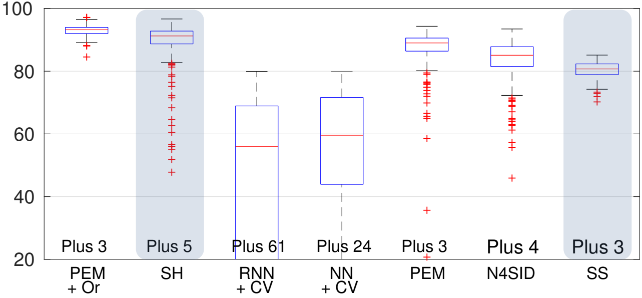


	Median of $\mathcal{F}(\hat{h})$
PEM+Or	93.2
PEM	89.0
N4SID	85.1

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

Impulse Response Fit:

$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\|h_{ij} - \hat{h}_{ij}\|}{\|h_{ij}\|}\right)$$

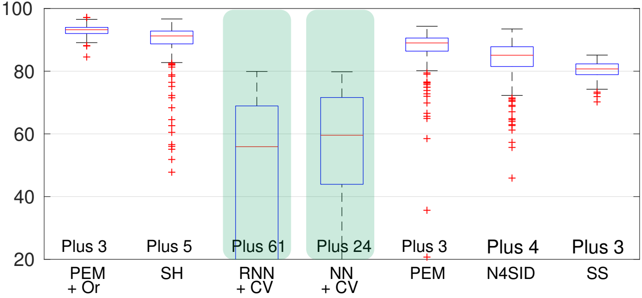


	Median of $\mathcal{F}(\hat{h})$
PEM+Or	93.2
SH	91.2
PEM	89.0
N4SID	85.1
SS	80.7

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

Impulse Response Fit:

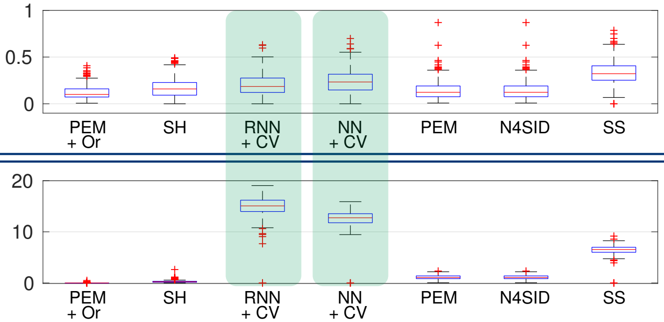
$$\mathcal{F}(\hat{h}) = \frac{100}{pm} \sum_{i=1}^{p} \sum_{j=1}^{m} \left(1 - \frac{\| h_{ij} - \hat{h}_{ij} \|}{\| h_{ij} \|} \right)$$



	Median of $\mathcal{F}(\hat{h})$
PEM+Or	93.2
SH	91.2
RNN+CV	55.9
NN+CV	59.6
PEM	89.0
N4SID	85.1
SS	80.7

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

$$\Delta_{signal}(\hat{h}) \quad := \quad \sum_{i=1}^{n} |s_i(h) \; - \; s_i(\hat{h})|$$

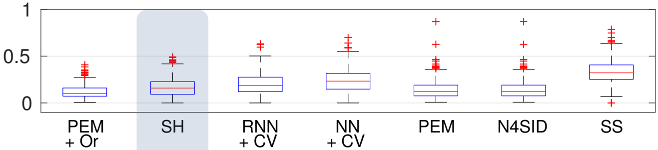


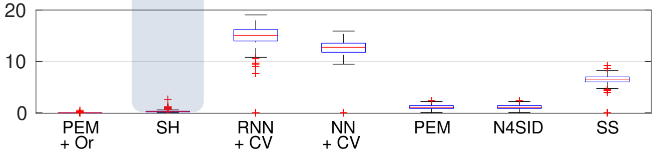
$$\Delta_{noise}(\hat{h}) \quad := \quad \sum_{i=n+1}^{r} |s_i(h) \; - \; s_i(\hat{h})|$$

Experimental Results

Setup: Fixed SIMO systems, 1 input/3 output, 200 MC runs, $N = 500$

$$\Delta_{signal}(\hat{h}) := \sum_{i=1}^{n} |s_i(h) - s_i(\hat{h})|$$





$$\Delta_{noise}(\hat{h}) := \sum_{i=n+1}^{r} |s_i(h) - s_i(\hat{h})|$$

Conclusion

- We have derived a new Gaussian prior to jointly enforce stability and low-complexity

Conclusion

- We have derived a new Gaussian prior to jointly enforce stability and low-complexity
- We have designed an iterative estimation algorithm

Conclusion

- We have derived a new Gaussian prior to jointly enforce stability and low-complexity
- We have designed an iterative estimation algorithm
- Numerical experiments show that our algorithm outperforms

Conclusion

- We have derived a new Gaussian prior to jointly enforce stability and low-complexity
- We have designed an iterative estimation algorithm
- Numerical experiments show that our algorithm outperforms
  - Non-parametric Bayesian method using only Stable-Spline prior (on MIMO systems)

Conclusion

- We have derived a new Gaussian prior to jointly enforce stability and low-complexity
- We have designed an iterative estimation algorithm
- Numerical experiments show that our algorithm outperforms
  - Non-parametric Bayesian method using only Stable-Spline prior (on MIMO systems)
  - Classical and state-of-the-art parametric methods (especially in presence of few data)

Thanks for your attention

References

H. Hjalmarsson, J. S. Welsh, and C. R. Rojas. "entification of Box-Jenkins models using structured ARX models and nuclear norm relaxation". In: *Proc. of SYSID.* 2012, pp. 322–327.

L. Ljung. *System Identification - Theory for the User*. 2nd. Upper Saddle River, N.J.: Prentice-Hall, 1999.

K. Mohan and M. Fazel. "Reweighted nuclear norm minimization with application to system identification". In: *American Control Conference (ACC), 2010*. IEEE. 2010, pp. 2953–2959.

G. Pillonetto and G. De Nicolao. "A new kernel-based approach for linear system identification". In: *Automatica* 46.1 (2010), pp. 81–93.

Publications

G. Prando and A. Chiuso. "Model reduction for linear Bayesian System Identification". In: *Proc. of IEEE CDC*. 2015.

G. Prando, A. Chiuso, and G. Pillonetto. "Bayesian and regularization approaches to multivariable linear system identification: the role of rank penalties". In: *Proc. of IEEE CDC*. Los Angeles, 2014.

G. Prando, A. Chiuso, and G. Pillonetto. "Maximum Entropy vector kernels for MIMO system identification". In: *Automatica* 79 (2017), pp. 326 –339.

G. Prando, G. Pillonetto, and A. Chiuso. "The role of rank penalties in linear system identification". In: *Proc. of 17th IFAC Symposium on System Identification, SYSID, Beijing*. 2015.

G. Prando et al. "Classical vs. Bayesian methods for linear system identification: point estimators and confidence sets". In: *Proc. of ECC*. 2016.

G. Prando, D. Romeres, and A. Chiuso. "Online Identification of Time-Varying Systems: a Bayesian approach". In: *Proc. of IEEE CDC*. 2016.

D. Romeres et al. "On-line Bayesian System Identification". In: *Proc. of ECC*. 2016.