Kernel manifold regression for the coupled electric drives dataset

Mirko Mazzoleni

Matteo Scandella

Fabio Previdi

mirko.mazzoleni@unibg.it
matteo.scandella@unibg.it
fabio.previdi@unibg.it

11 April 2018

Outline

1. Introduction and motivation

2. A new framework for nonparametric system identification

3. Application to the coupled electric drive dataset

4. Conclusions and future developments

Outline

1. Introduction and motivation

2. A new framework for nonparametric system identification

3. Application to the coupled electric drive dataset

4. Conclusions and future developments

Introduction and motivation

Data

Distance

Model

Introduction and motivation

Data

Distance

Model

Model definition

- Consider the **NARX system**: $\mathcal{S}: y(t+1) = g\big(x_u(t), x_y(t)\big) + e(t)$, where:
  - ✓ $y(t) \in \mathbb{R}$ is the system output
  - ✓ $g(t)$ is a nonlinear function
  - ✓ $x_u(t) = [u(t), \dots, u(t-p+1)]^T \in \mathbb{R}^{p\times 1}$ is a regressor vector of past p inputs
  - ✓ $x_y(t) = [y(t), \dots, y(t-q+1)]^T \in \mathbb{R}^{q\times 1}$ is a regressor vector of past q outputs
  - ✓ $x\,(t) = \big[x_y(t), x_y(t)\big]^T \in \mathbb{R}^{p+q\times 1}$
  - ✓ $e(t) \in \mathbb{R}$ is an additive white noise

Reproducing Kernel Hilbert Spaces (RKHS)

- An RKHS is a Hilbert space $\mathcal{H}$ such that:

  a. *Its elements are functions* $u: \Omega \to \mathbb{R}$ *, where* Ω *is a generic set*

  b. $\forall x \in \Omega, L_x: \mathcal{H} \to \mathbb{R}$ *is a continuous linear functional*
  $u \to u(x)$

- Riesz's representation theorem $\rightarrow$ $\boxed{\exists\, r_x \in \mathcal{H} \text{ s.t. } L_x(u) = u(x) = \langle u, r_x \rangle}$

- The function $r_x(\cdot)$ is called the representer of evaluation in x

Reproducing Kernel Hilbert Spaces (RKHS)

- The reproducing kernel is defined as: $K(x,z) = \langle r_x, r_z \rangle, \quad K = \Omega \times \Omega \to \mathbb{R}$

  a. *Symmetric:* $K(x,z) = K(z,x)$

  b. *Semidefinite positive* $\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0 \quad \forall n, c_i \in \mathbb{R}, \forall x_i \in \Omega$

- Moore-Aronszajn theorem → A RKHS defines a corresponding reproducing kernel. Conversely, a reproducing kernel defines a unique RKHS

Examples of kernels

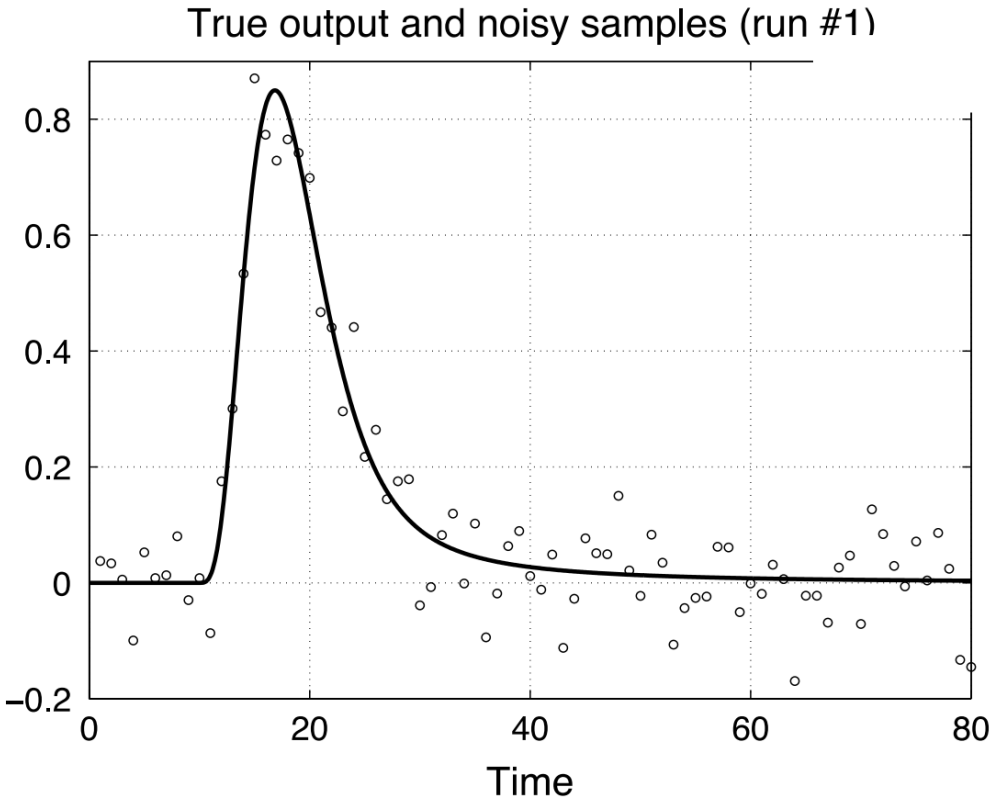
- Constant kernel: $K(x, z) = 1$
- Linear kernel: $K(x, z) = x \cdot z$
- Polynomial kernel: $K(x, z) = (x \cdot z + 1)^d$
- Gaussian kernel: $K(x, z) = e^{-\frac{\|x-z\|^2}{2\sigma^2}}$

Kernel composition theorem:

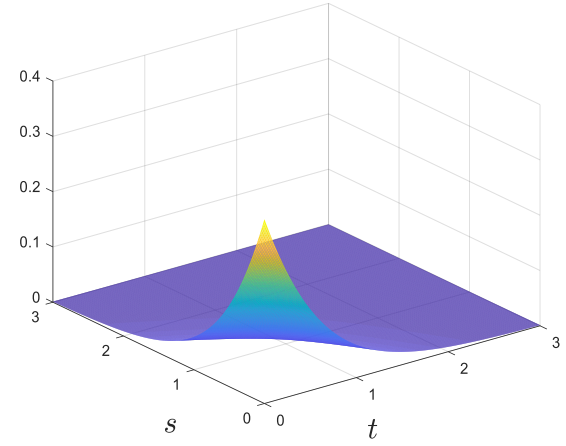
- A linear combination of valid kernel functions is a valid kernel function
- The space induced by this kernel is the span of the spaces induced by the single ones

$$H = \bigoplus_i H_i$$

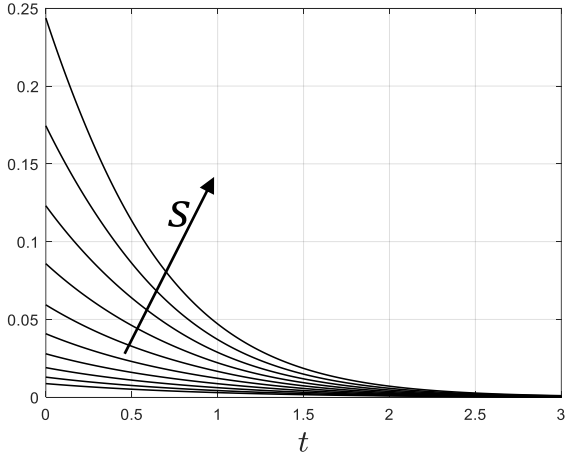
Kernel methods in system identification



Stable spline kernel



Representers



- *Pillonetto, Gianluigi and Giuseppe De Nicolao. "A new kernel-based approach for linear system identification." Automatica 46 (2010): 81-93.*
- *Pillonetto, Gianluigi et al. "A New Kernel-Based Approach for NonlinearSystem Identification." IEEE Transactions on Automatic Control 56 (2011): 2825-2840.*

Introduction and motivation

Data

Model

Distance

Nonparametric learning

- Consider the variational formulation:

$$\hat{g} = \arg\min_{g \in \mathcal{H}} \sum_{i=1}^{N} \left(y_i - g(x_i)\right)^2 + \lambda_T \cdot \|g\|_{\mathcal{H}}^2 \qquad y_i = y(t_i);\ \ x_i = x(t_i)$$

- Tikhonov regularization: $\|g\|^2_{\mathcal{H}}$ penalizes the norm of the fitted function
- The minimization problem is on the RKHS space $\mathcal{H}$ → infinite number of parameters!

Representer theorem

- The minimizer of the variational problem is given by:

$$\hat{g}(x) = \sum_{i=1}^{N} c_i K(x, x_i) = \sum_{i=1}^{N} c_i r_{x_i}(x)$$

For some N-tuple
$c = [c_1, c_2, \dots, c_N]^T \in \mathbb{R}^{N \times 1}$

- Linear combination of the representers of the training points x_i, evaluated in the point x
- The solution is expressed as combination of «basis functions» which properties are determined by $\mathcal{H}$

Nonparametric learning – Solution

- Using the representer theorem it possible to express the variational problem as:

$$\hat{c} = \arg\min_{c \in \mathbb{R}^N} \|Y - \mathcal{K}c\|_2^2 + \lambda_T \cdot c^T \mathcal{K} c$$

- $Y \in \mathbb{R}^{N \times 1}$: vector of observations
- $\mathcal{K} \in \mathbb{R}^{NxN}$: semidefinite positive and symmetric matrix, s.t. $\mathcal{K}_{ij} = K(x_i, x_j)$

- Since the expression is quadratic in c we have the closed-form solution:

$$\boxed{(\mathcal{K} + \lambda_T \cdot I_N) \cdot \hat{c} = Y}$$

Outline

1. Introduction and motivation

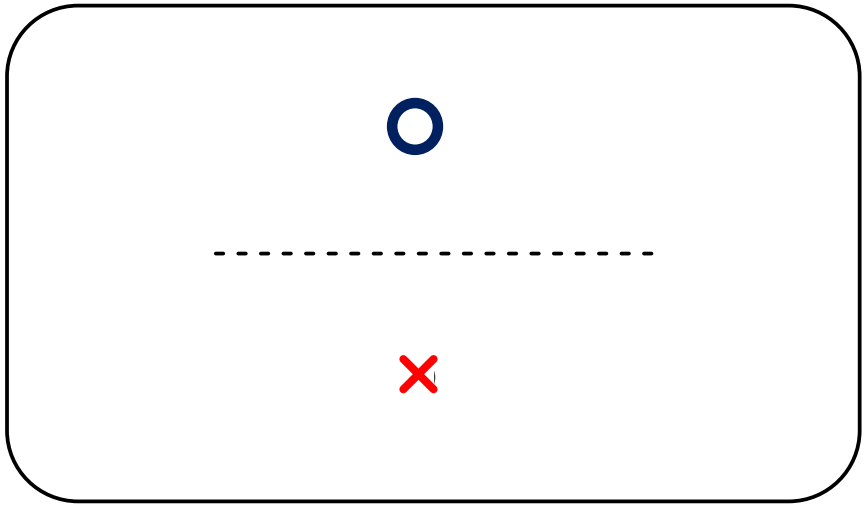
2. A new framework for nonparametric system identification

3. Application to the coupled electric drive dataset

4. Conclusions and future developments

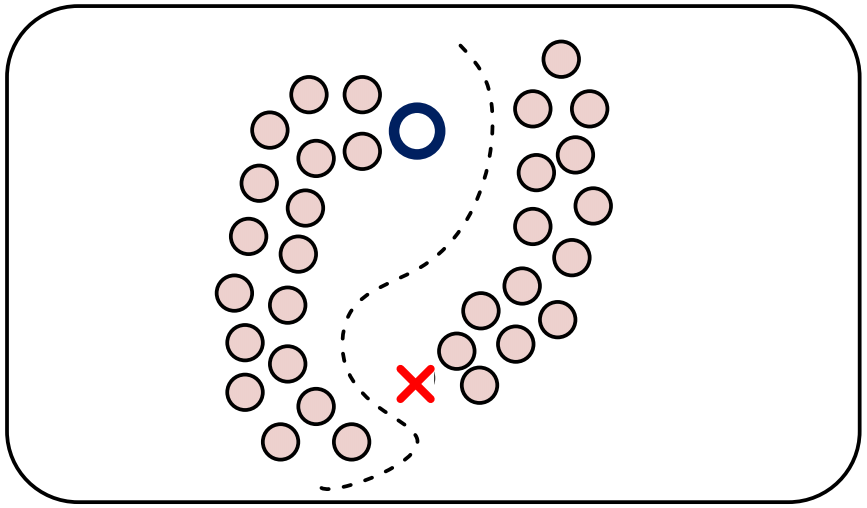
Manifold learning (static systems)

- Suppose that the regressors' belong to a **manifold** in the regressors' space
- The position of the regressors adds prior information
- How to incorporate this information in a learning framework?



Manifold learning (static systems)

- Suppose that the regressors' belong to a **manifold** in the regressors' space
- The position of the regressors adds prior information
- How to incorporate this information in a learning framework?



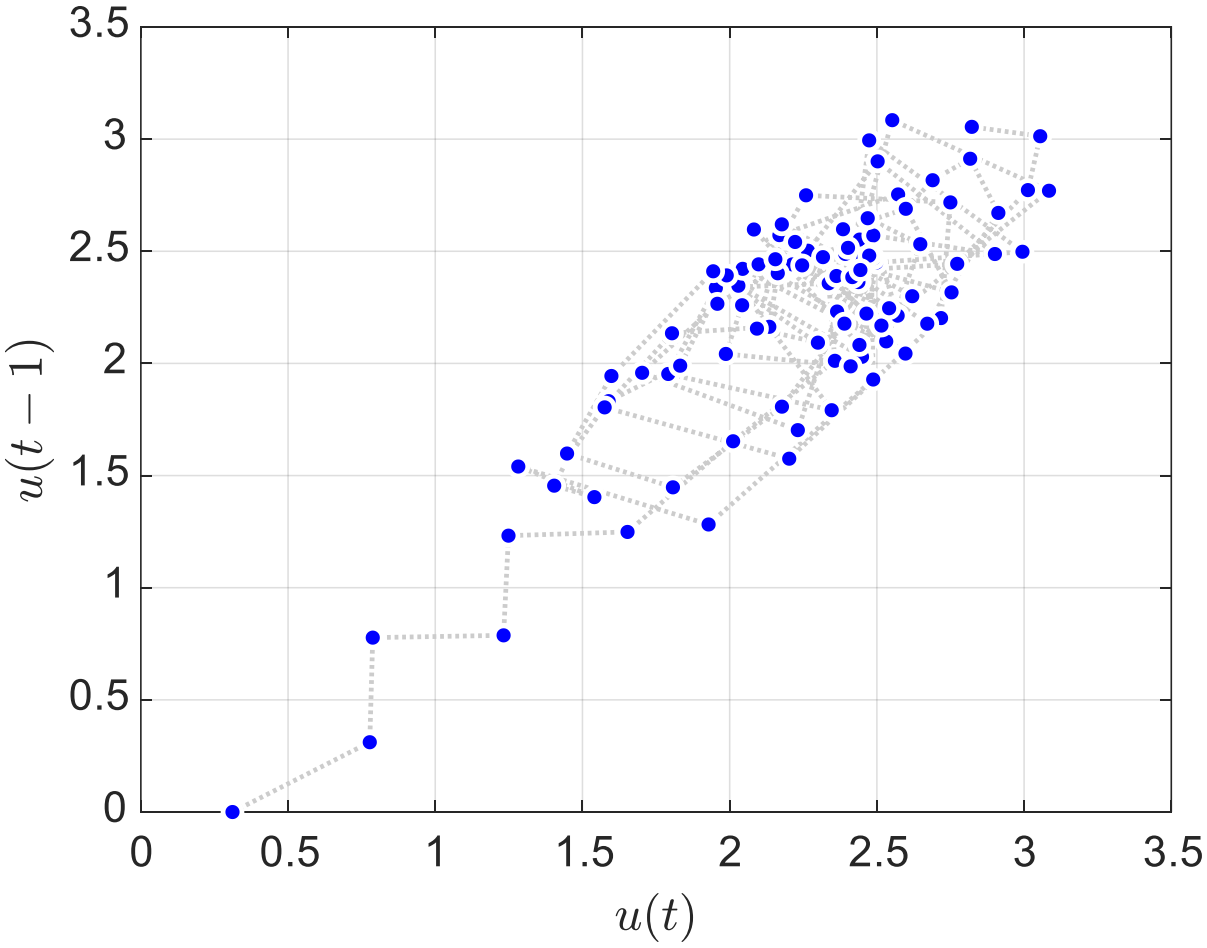
Incorporating the manifold information

Semi-supervised smoothness assumption

If two regressors $\boldsymbol{x}(\boldsymbol{i})$ *and* $\boldsymbol{x}(\boldsymbol{j})$ *in a high-density region are close, then so should be their corresponding outputs* $\boldsymbol{y}(\boldsymbol{i})$ *and* $\boldsymbol{y}(\boldsymbol{j})$

Link to dynamical systems

- In dynamical systems, regressors can be strongly correlated
- It is meaningful to think that they lie on a manifold of the regressors' space
- PCA reveals how 91% of variance explained by one component



Manifold regularization

- One way to enforce the smoothness assumption is to minimize the quantity:

$$S_g = \int_{\mathcal{G}} \|\nabla \cdot g\|^2 \, dp(x) = \int_{\mathcal{G}} g \cdot \Delta \cdot g \, dp(x)$$

- $p(x)$: probability distribution of the regressors, ∇: Gradient, Δ: Laplacian-Beltrami operators
- Minimizing S_g means minimizing the gradient of the function
- The term can rarely be computed since $p(x)$ and $\mathcal{G}$ are unknown

Manifold regularization

- The term S_g can be modeled using the *regressor graph*, encoding connections and distasnces between points:
  - ✓ with the regressors as its vertices
  - ✓ the weights on the edges are defined as: $w_{i,j} = e^{-\frac{\|x(i)-x(j)\|^2}{2\sigma_e^2}}$
- Considering the Laplacian matrix $L = D - W \in \mathbb{R}^{N \times N}$, where:
  - ✓ $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix $D_{ii} = \sum_{j=1}^{N} w_{i,j}$; $W \in \mathbb{R}^{N \times N}$ contains the $w_{i,j}$

- σ_e is a tuning parameter
- Higher value ➔ similar regressors

Manifold regularization

- We have therefore that:

$$S_g = \int_{\mathcal{G}} \|\nabla \cdot g\|^2 \, dp(x) = \int_{\mathcal{G}} g \cdot \Delta \cdot g \, dp(x) \approx F^T \cdot L \cdot F$$

- The vector $F = \left[g\left(x(1)\right), \cdots, g\left(x(N)\right)\right]^T \in \mathbb{R}^{N \times 1}$ contains the ***noiseless*** outputs

- To compute the approximation, only the regressors are needed

Manifold regularization

- We can than enforce the smoothness assumption by adding a new regularization term:

$$\hat{g} = \arg\min_{g \in \mathcal{H}} \sum_{i=1}^{N} \left(y(t) - g\left(x(t)\right)\right)^2 + \lambda_T \cdot \|g\|_{\mathcal{H}}^2 + \lambda_M \cdot F^T \cdot L \cdot F$$

- The representer theorem still holds for this cost function
- The solution to the minimization problem admits a closed-form solution:

$$\boxed{(\mathcal{K} + \lambda_T \cdot I_N + \lambda_M \cdot L \cdot \mathcal{K}) \cdot \hat{c} = Y} \quad \Longrightarrow \quad \boxed{A \cdot \hat{c} = Y}$$

Outline

1. Introduction and motivation

2. A new framework for nonparametric system identification

3. Application to the coupled electric drive dataset

4. Conclusions and future developments

Data

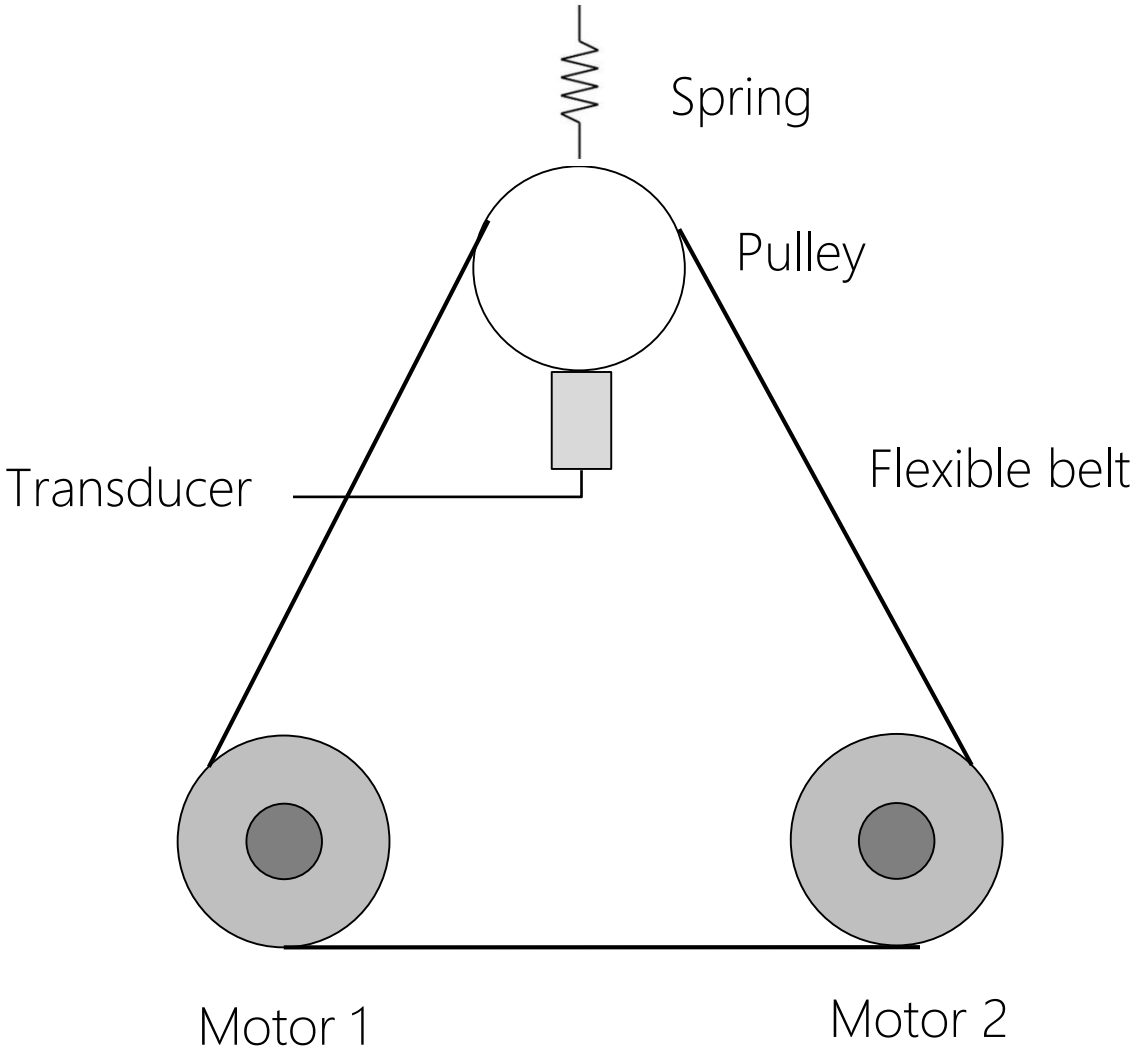
Distance

Model

Application to coupled electric drive dataset

Results

- Application to the coupled electric drives
- Input: motor voltage Uniformely distributed dataset $N = 500$
- Output: pulley speed
- Ts = 20ms



Results

- The method has been applied to the coupled electric drives dataset

- The following kernel was employed:

$$K(x,z) = \eta_{nl} \cdot e^{-\frac{\|x-z\|^2}{\sigma^2}} + \eta_l \cdot x^T z + \eta_c$$

Hyperparameters $\psi = [\lambda_T, \lambda_M, \eta_{nl}, \eta_l, \eta_c, \sigma, \sigma_e]$

Results

- The hyperparameters have been estimated using Generalized Cross Validation (GCV):

$$\hat{\psi}_{p,q} = \underset{\psi}{\operatorname{argmin}} \frac{N}{\left(N-\nu(\psi)\right)^2} \cdot \left\|Y - \hat{Y}\right\|_2^2 = \underset{\psi}{\operatorname{argmin}} J_{p,q}(\psi)$$

- Where $\nu(\psi)$ are the degree of freedom of the model:

$$\hat{Y} = \mathcal{K} \cdot \hat{c} = \mathcal{K} \cdot A^{-1} \cdot Y = S \cdot Y \quad \Rightarrow \quad \nu(\psi) = tr[S]$$

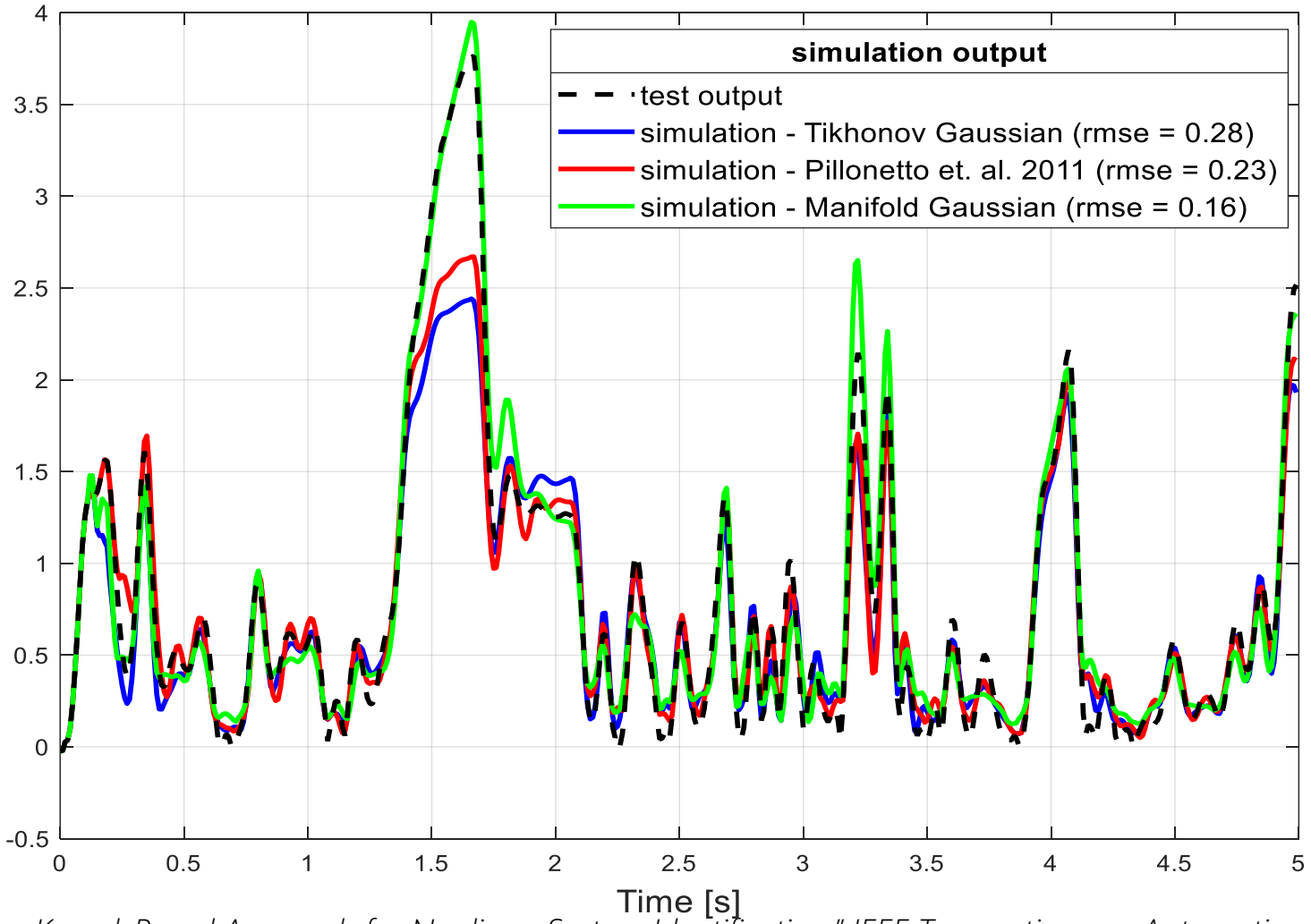
- The order of the system is estimated via a grid search as:

$$(\hat{p}, \hat{q}) = \underset{(p,q)}{\operatorname{argmin}} J_{p,q}\left(\hat{\psi}_{p,q}\right)$$

Application to coupled electric drive dataset

Results

- Simulation results improve over existing approaches
- All kernels equipped with linear and constant terms



- *Gianluigi Pillonetto, Minh Ha Quang, Alessandro Chiuso. "A New Kernel-Based Approach for NonlinearSystem Identification." IEEE Transactions on Automatic Control 56 (2011): 2825-2840*

Outline

1. Introduction and motivation

2. A new framework for nonparametric system identification

3. Application to the coupled electric drive dataset

4. Conclusions and future developments

Conclusions

- A new paradigm for nonparametric learning of nonlinear system was presented
- The idea leverages the concept of manifold regularization
- Future challenges (we are working on it already):
  - ✓ **Bayesian derivation** of the manifold regularization approach
  - ✓ Application to wider range of systems and nonlinearities

References

- M. Mazzoleni, S. Formentin, M. Scandella, F. Previdi «*Semi-supervised learning of dynamical systems: a preliminary study*.» 16th European Control Conference (ECC), Limassol, Cyprus, 2018. In press.

- M. Mazzoleni, M. Scandella, S. Formentin, F. Previdi «*Identification of nonlinear dynamical system with synthetic data: a preliminary investigation*.» 18th IFAC Symposium on System Identification, Stockholm, Sweden, 2018. In press.

THANKS FOR YOUR ATTENTION