# From Volterra series through block-oriented approach to Volterra series

**Pawel Wachel and Przemyslaw Sliwinski**

Department of Control Systems and Mechatronics
Wroclaw University of Technology

European Research Network on System Identification (ERNSI)
Varberg 2015

# Outline

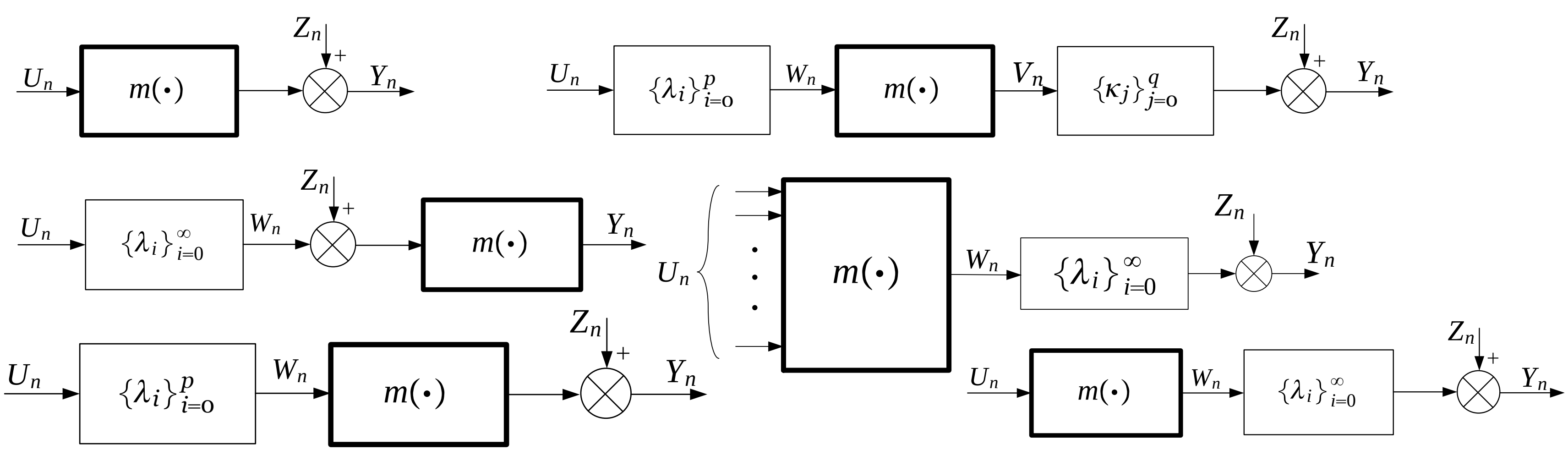
- Introduction – Volterra series and block oriented models
- Nonparametric kernel methods – basic concepts
- Nonparametric kernel regression in application to system identification
- Mixed (parametric–nonparametric) system identification
- Aggregative modeling – dictionary approach
- Application of aggregative approach to Volterra series modeling of LNL system

# Volterra vs. block–oriented approach

## Classic Volterra approach

**Volterra series** + **Classic identification methods** = **Curse of dimensionality problems...**

## Block–oriented „alternatives”



# General block–oriented identification setup

**Input signal**

*i.i.d. random sequence*

**Nonlinearity**

*Nonparametric prior knowledge*

**Dynamic subsystem(s)**

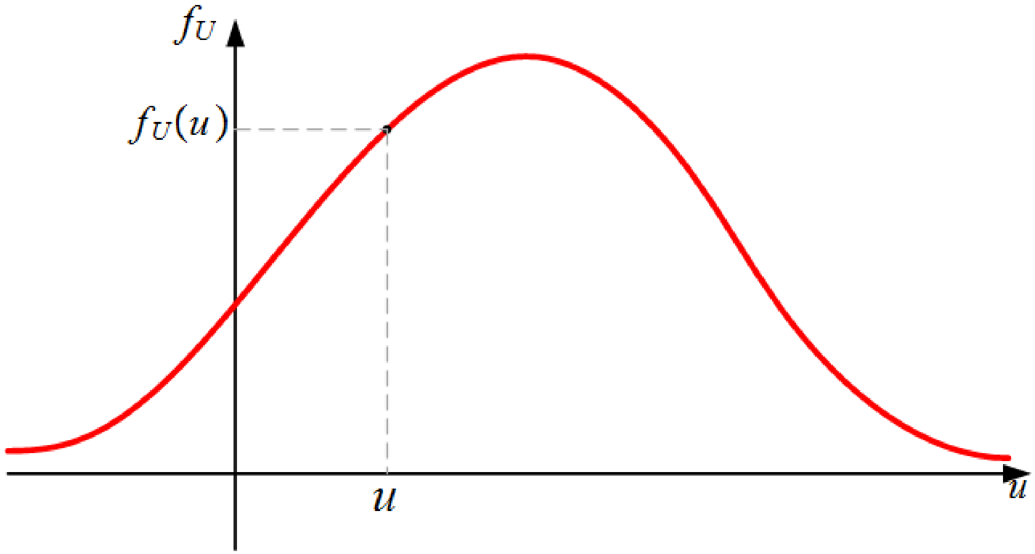
*Asymptotic stability assumed*

# Nonparametric kernel density estimation

- Let us consider the following estimation problem: Given the set of *i.i.d.* measurements

$$U_1, U_2, \ldots, U_N$$

with the **unknown** probability density function $f_U(\cdot)$

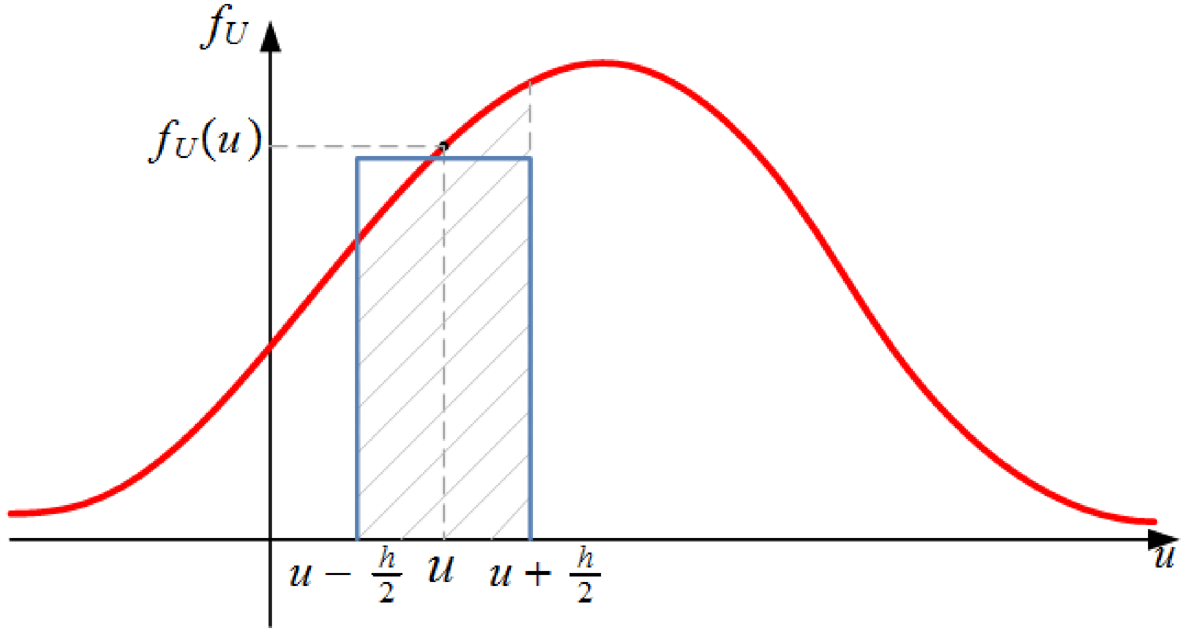



find (estimate) $f_U(\cdot)$ in some point $u$ from the set $U_1, U_2, \ldots, U_N$

# Nonparametric kernel density estimation

- The main idea behind the considered estimate is based on the observation that for any $u$ and relatively small constant $h$:

$$f_U(u) \approx \frac{1}{h} \int_{u-h/2}^{u+h/2} f_U(v)dv.$$

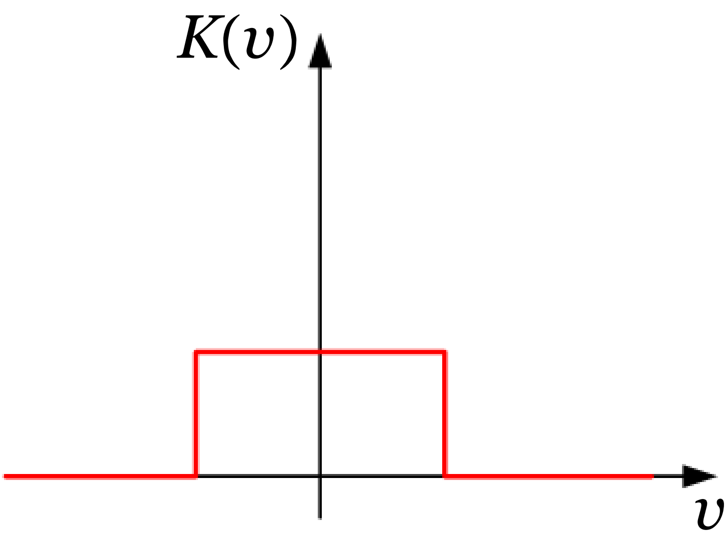


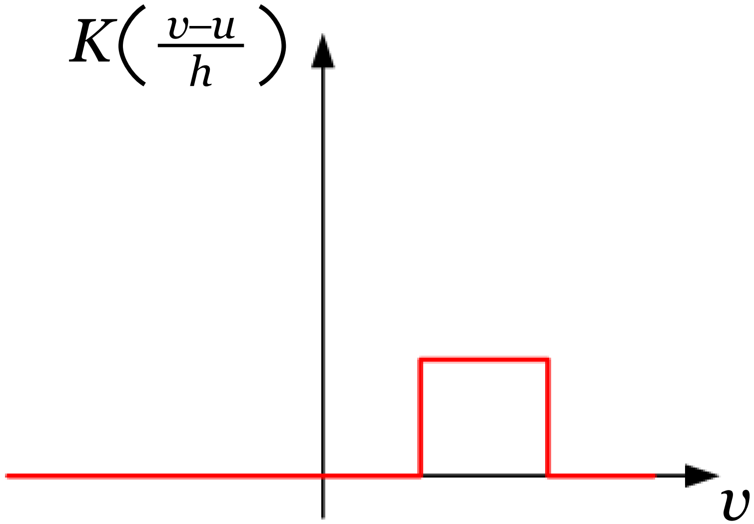
# Nonparametric kernel density estimation

- Let now $K(\cdot)$ be a box kernel function of the form

$$K(v) = \begin{cases} 1 & \text{for} \quad v \in \left\langle -\frac{1}{2}; \frac{1}{2} \right\rangle \\ 0 & \text{for} \quad v \notin \left\langle -\frac{1}{2}; \frac{1}{2} \right\rangle \end{cases}$$

**How the kernel works?**



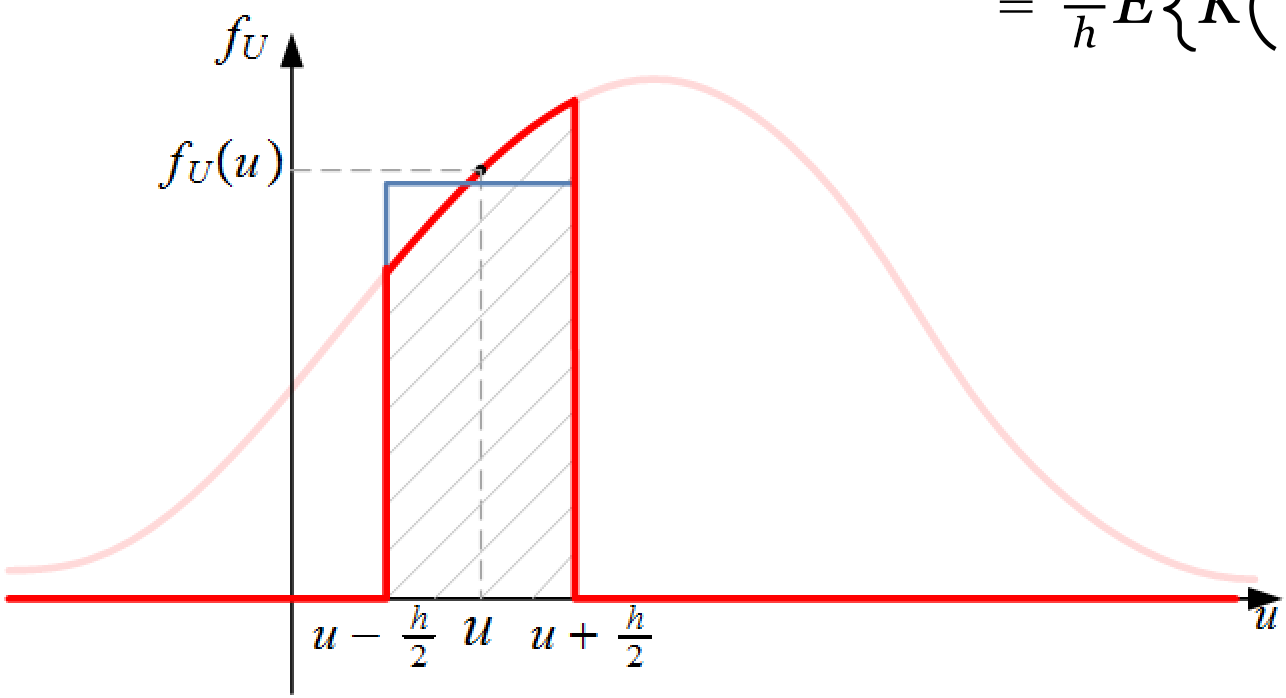



# Nonparametric density estimation

- Using the kernel function we obtain further

$$f_U(u) \approx \frac{1}{h} \int_{u-h/2}^{u+h/2} f_U(v)dv = \frac{1}{h} \int_{-\infty}^{+\infty} f_U(v) K\left(\frac{v-u}{h}\right) dv$$

$$= \frac{1}{h} E\left\{K\left(\frac{U_1-u}{h}\right)\right\}$$



# Nonparametric density estimation

- Finally, on the basis of

$$f_U(u) \approx \tfrac{1}{h} E\left\{K\left(\tfrac{U_1-u}{h}\right)\right\}$$

we obtain the following Kernel Density Estimate (Rosenblatt, Parzen)

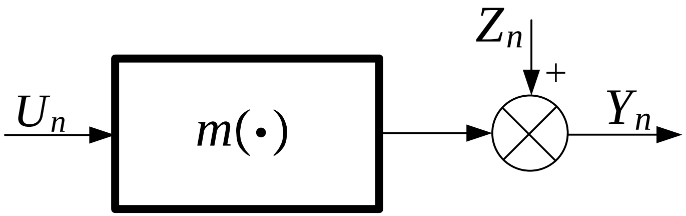
$$\hat{f}_U(u) = \tfrac{1}{Nh} \sum_{i=1}^{N} K\left(\tfrac{U_i-u}{h}\right)$$

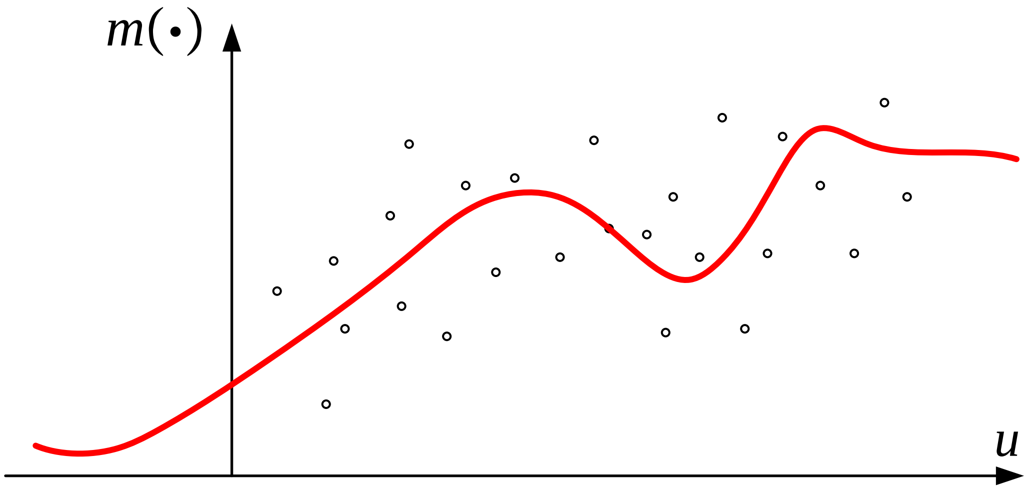
(which is a sum of scaled and translated kernel functions)

# Nonparametric kernel identification

- Consider a static nonlinear system described by the formula

$$Y_n = m(U_n) + Z_n$$



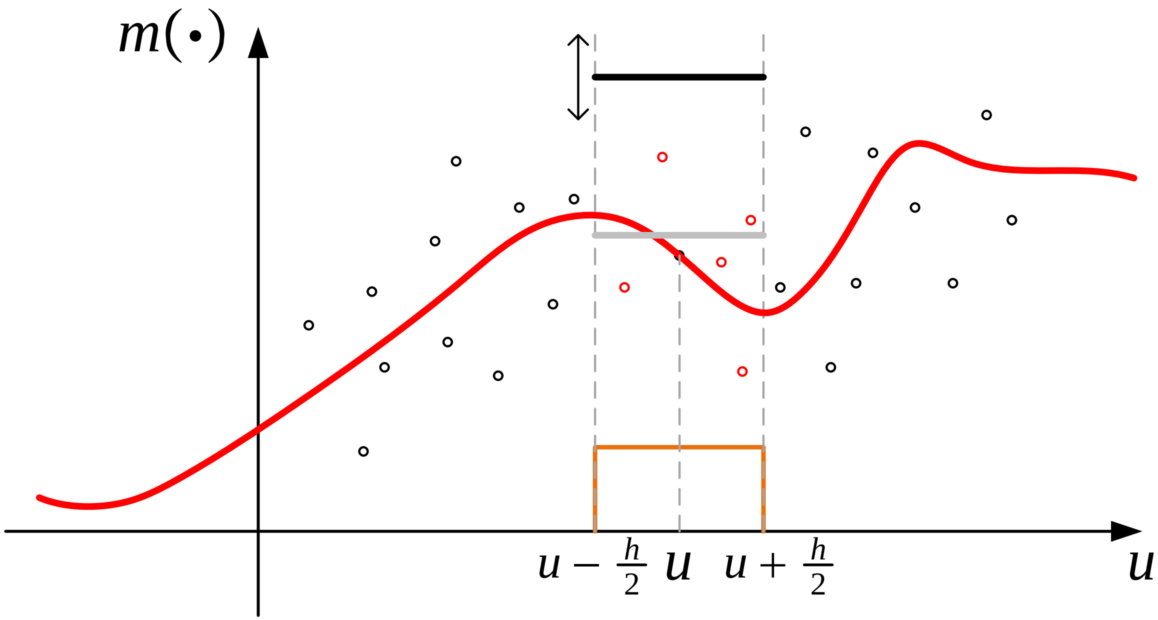



# Nonparametric kernel identification

- Consider the following local least squares quality function

$$\tilde{Q}(c;u) = \sum_{i=1}^{N}[Y_i - c]^2 K\left(\frac{U_i - u}{h}\right)$$

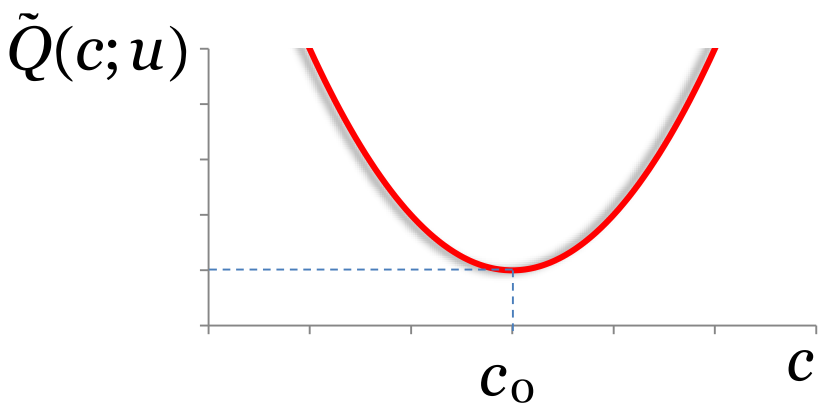
- Goal: find $c_0$ such that $c_0 = \operatorname{argmin}_{c\in\mathbb{R}} \tilde{Q}(c;u)$



# Nonparametric kernel identification

$$\tilde{Q}(c;u) = \sum_{i=1}^{N}[Y_i - c]^2 K\left(\frac{U_i - u}{h}\right)$$

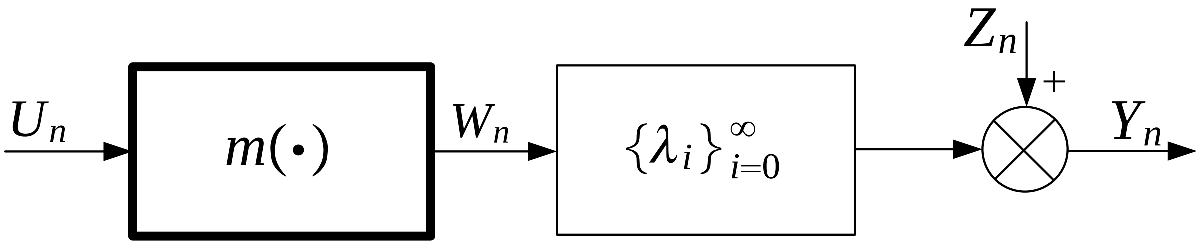
$\tilde{Q}(c;u)$ is a (non-negative) **quadratic** function of variable $c$.




**Solution**: Nadaraya–Watson kernel estimate

$$c_0 = \hat{m}(u) = \frac{\sum_{i=1}^{N} Y_i K\left(\frac{U_i - u}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{U_i - u}{h}\right)}$$

# Hammerstein system – nonlinearity estimation




$$W_n = m(U_n),$$

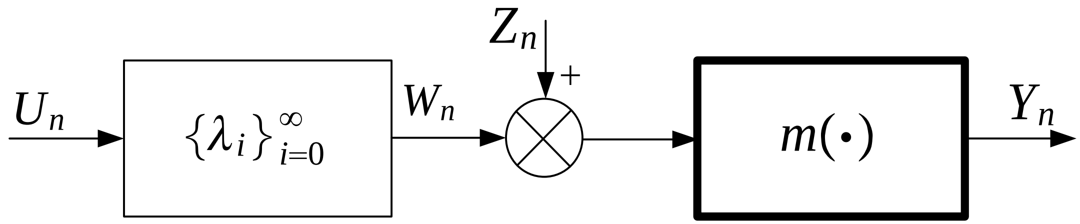
$$Y_n = \sum_{i=0}^{\infty} \lambda_i W_{n-i} + Z_n$$

$$\hat{\mu}(u) = \frac{\sum_{i=1}^{N} Y_i K\left(\frac{U_i - u}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{U_i - u}{h}\right)}$$

$\hat{\mu}(u) \to \alpha \cdot m(u) + \beta$, as $N \to \infty$ in probability

# Wiener system

Gaussian input assumed




$$W_n = \sum_{l=0}^{\infty} \lambda_l U_{n-l} + Z_n,$$

$$Y_n = m(W_n).$$

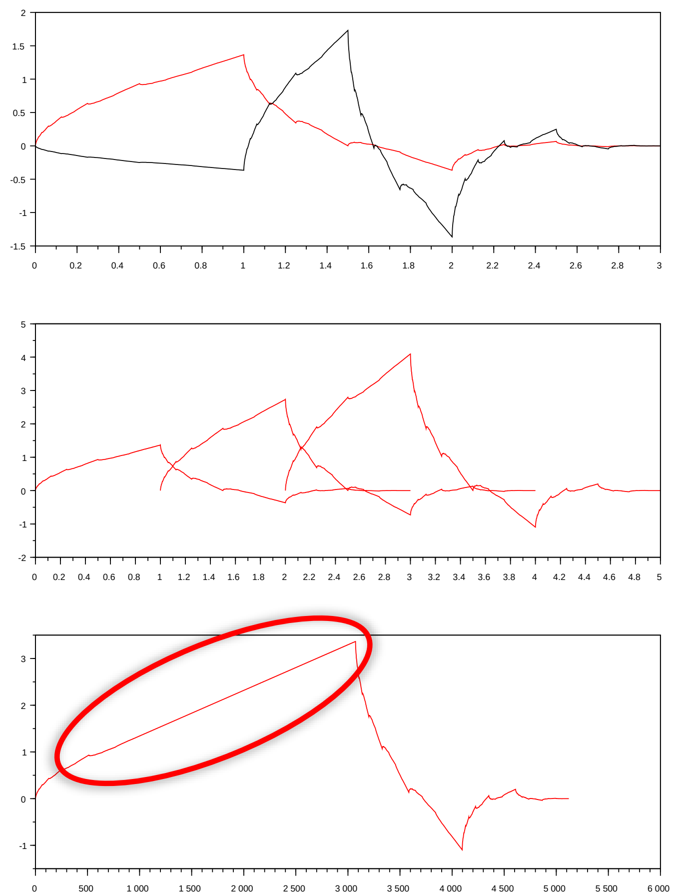
**Main concept**: Replace input measurements with output ones in the kernel estimate.

$$\hat{\mu}(y) = \frac{\sum_{i=1}^{N} U_i K\left(\frac{Y_i - y}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{Y_i - y}{h}\right)}$$

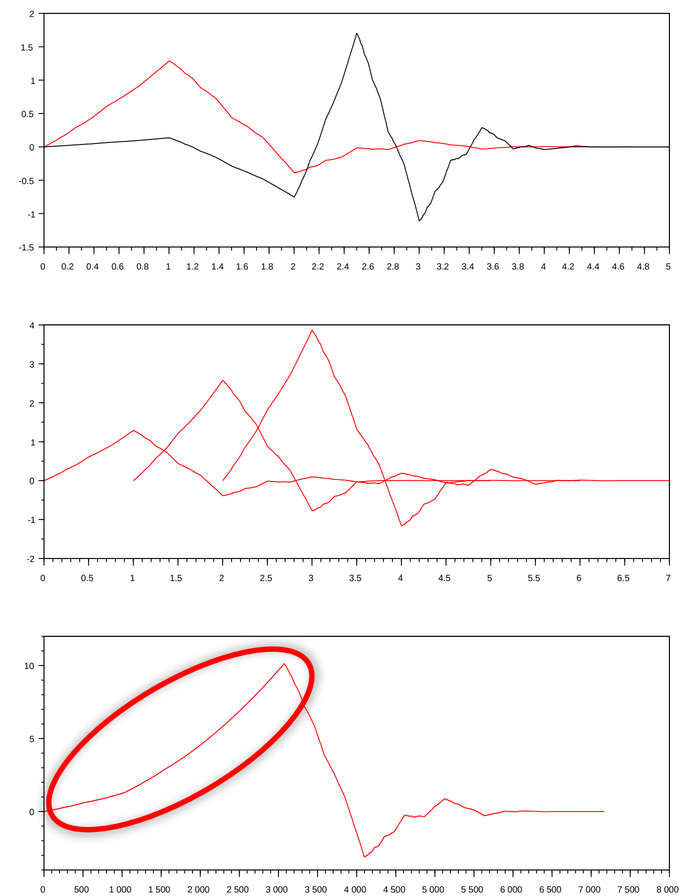
$\hat{\mu}(y) \to \alpha m^{-1}(y)$ as $N \to \infty$ in probability.

# Orthogonal algorithms and wavelets

## 2nd order Daubechies




## 3rd order Daubechies



# Orthogonal wavelet regression estimate

Orthogonal estimate of the nonlinearity

$$\hat{\mu}_K(u) = \frac{\sum_{i=1}^{N} Y_i \vartheta_K(u, U_i)}{\sum_{i=1}^{N} \vartheta_K(u, U_i)}$$

NOTE SIMILARITY!

$$\hat{\mu}(u) = \frac{\sum_{i=1}^{N} Y_i K\left(\frac{U_i - u}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{U_i - u}{h}\right)}$$

where

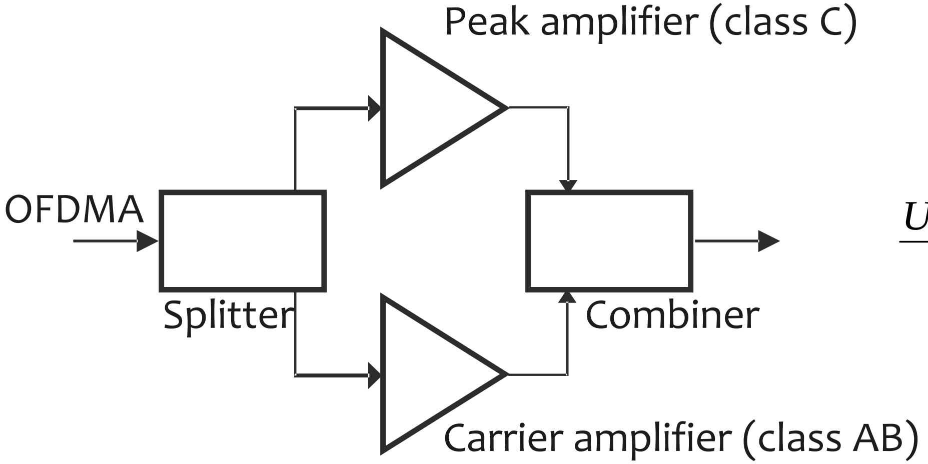
$$\vartheta_K(u, v) = \sum_{n=n_1}^{n_2} \varphi_{Kn}(u) \varphi_{Kn}(v)$$

and $\{\varphi_{Kn}(\cdot)\}$ is a family of orthogonal functions.

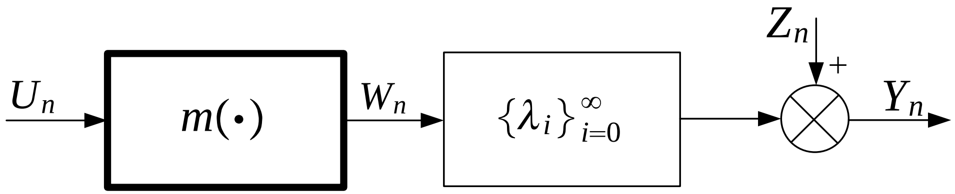
**Predistortion problem:** Estimate **an inverse** of the system's nonlinearity $m^{-1}(\cdot)$ and linearize system by replacing input $U_n$ with $m^{-1}(U_n)$.

**Doherty amplifier**

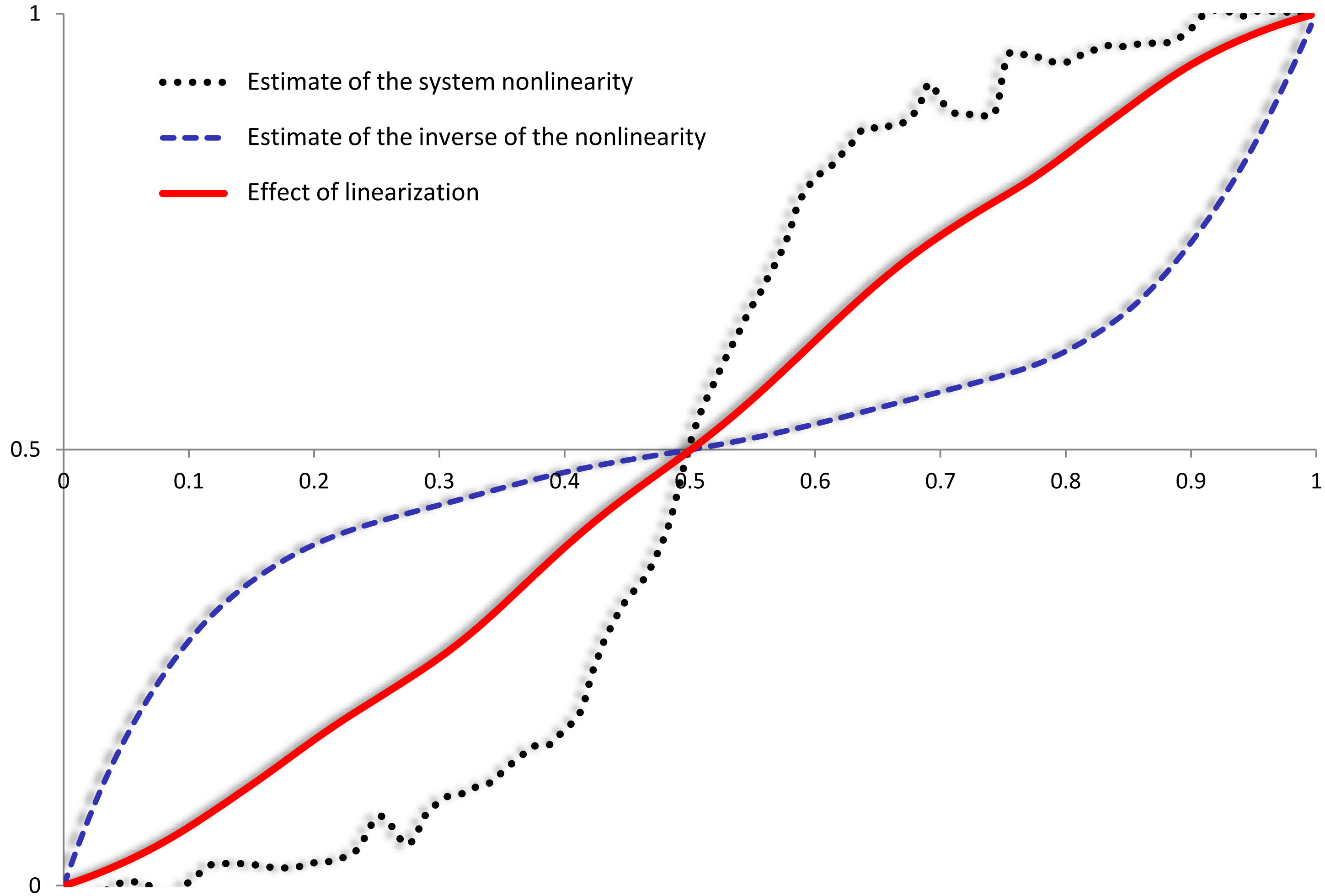
**Scheme**




**Hammerstein model**



# Predistortion – practical example

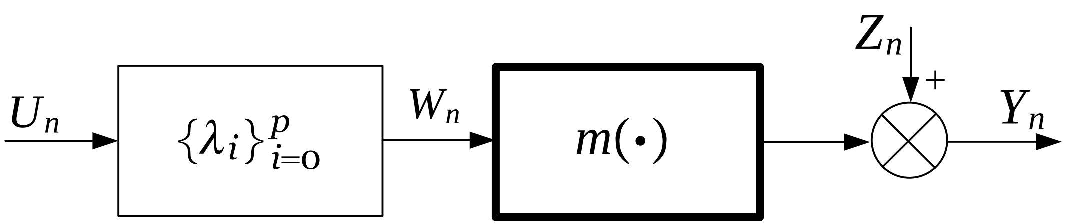


# Average Derivative Estimation – Wiener system

**Fact** (*chain rule*): Let $m(\cdot)$ be a differentiable function, then

$$\frac{d}{dx} m(\alpha x) = \alpha \left[ \frac{d}{dx} m(\alpha x) \right]$$

**Conclusion**: Derivative can **extract** factor $\alpha$ embedded in the argument of $m(\cdot)$.



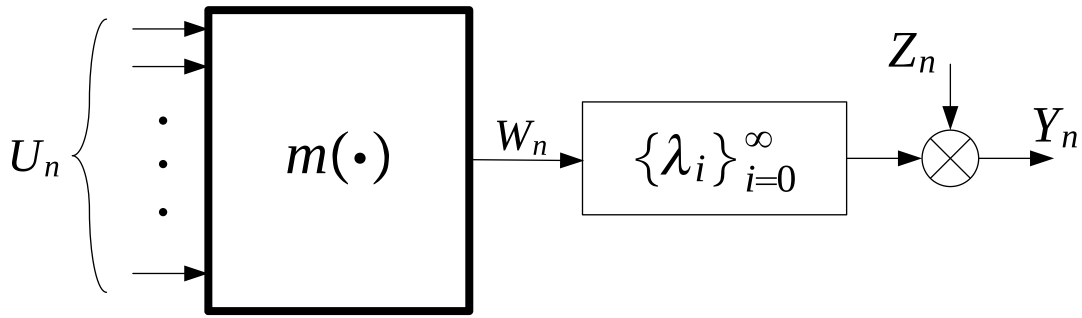

$$Y_n = m\left(\sum_{i=0}^{p} \lambda_i U_{n-i}\right) + Z_n$$

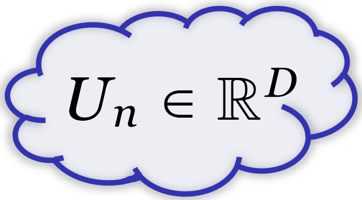
VECTOR NOTATION

$$Y_n = m(\boldsymbol{\lambda}^T \mathbf{U}_n) + Z_n$$

# Identification from structured input data

- Consider the following MISO Hammerstein system:

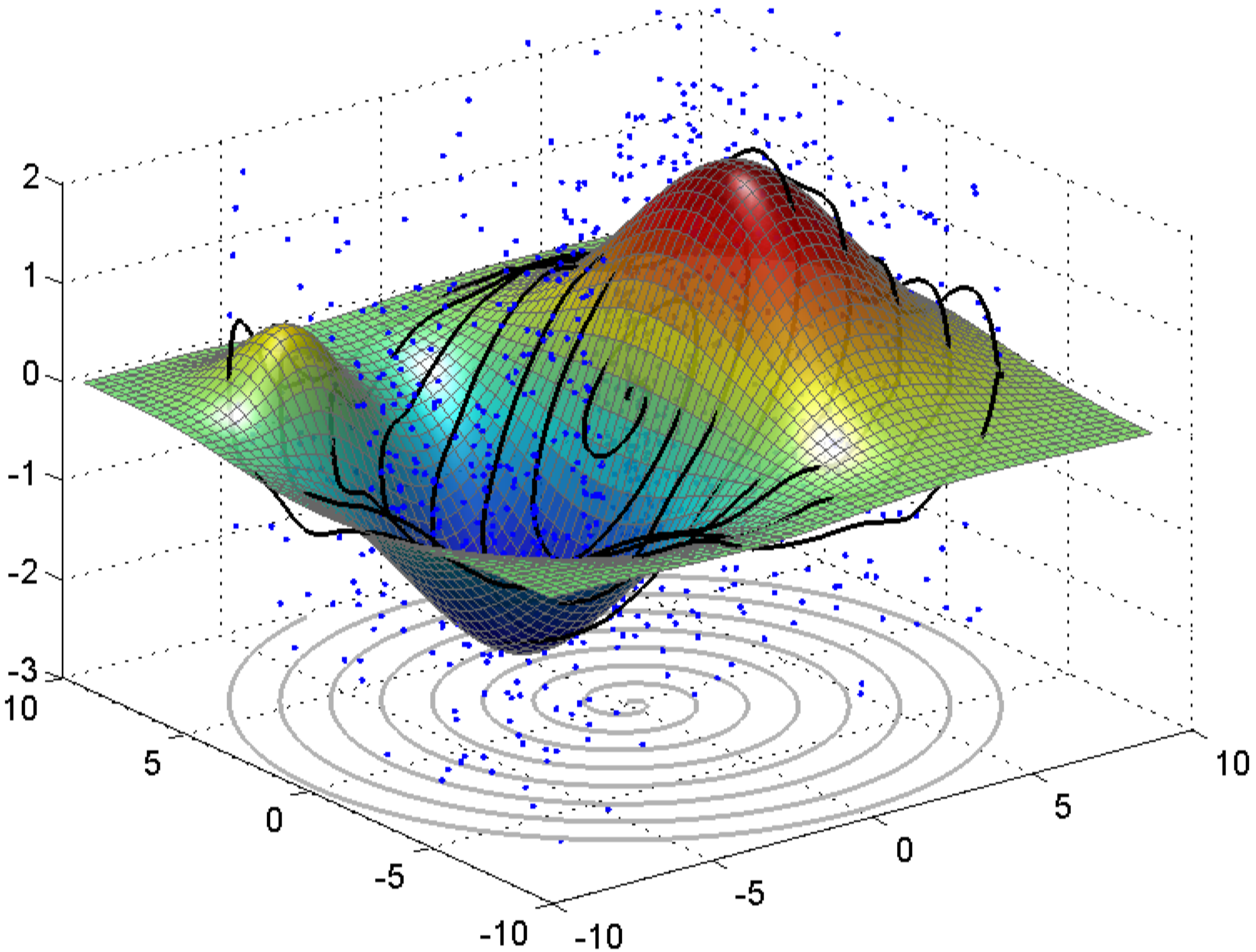


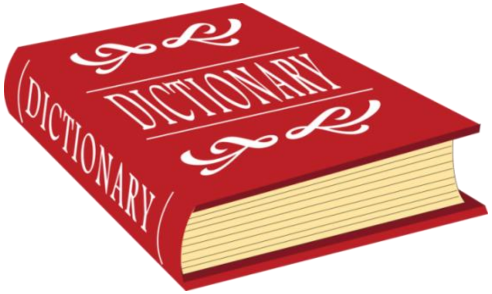

**Fact:** Convergence rate of the nonparametric estimators strongly depends on the number of system inputs.

**Assumption**: Input data are structured *i.e.* grouped on unknown $d$–dimensional manifold ($d < D$).

# Identification from structured input data

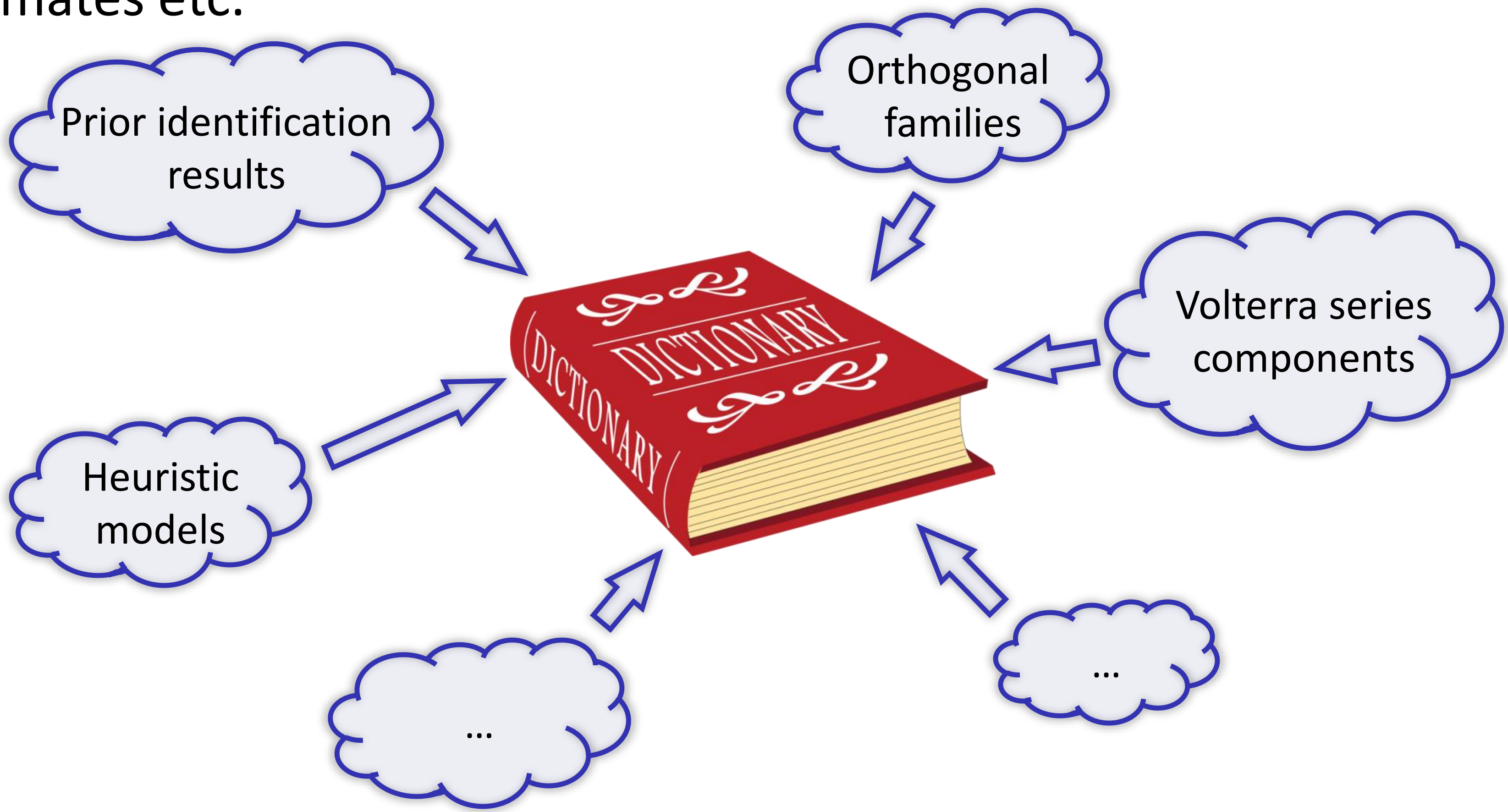


# AGGREGATIVE MODELING



# Dictionary modeling

The *a priori* knowledge about the system is **collected in a dictionary** in a form of the system components, models, pre-estimates etc.



# Aggregative modeling

- The class of considered SISO systems can be described by the general discrete-time input-output equation

$$Y_n = m(X_n, X_{n-1}, X_{n-2}, \ldots, X_{n-p}) + Z_n,$$

**Assumptions:**

- $\{X_n\}$ is a stationary sequence of inpendent random variables.
- $m(\mathbf{x})$ is **any** bounded function, such that $|m(\mathbf{x})| \leq L < \infty$
- $\{Z_n\}$ is a zero-mean i.i.d. random sequence with finite variance.

- There is a dictionary of maps,

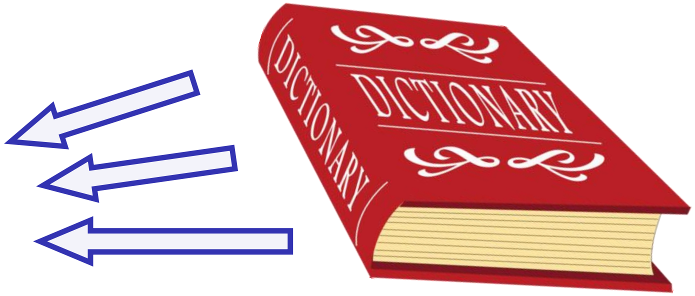
$$S = \{\bar{m}_i : \mathbb{R}^{p+1} \to [-L, L], i = 1, \ldots, D\}, D > 2$$

collected by a user to model the system.

# Aggregative modeling

The algorithm combines all the dictionary entries into a single one, referred to as the aggregated empirical model.

$$m(\mathbf{x};\hat{\boldsymbol{\alpha}}) = \sum_{i=1}^{D} \hat{\alpha}_i \bar{m}_i(\mathbf{x})$$




where

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{\alpha}} \hat{Q}(\boldsymbol{\alpha}), \text{ subject to } \|\boldsymbol{\alpha}\|_1 \leq 1,$$

and where

$$\hat{Q}(\boldsymbol{\alpha}) = \frac{1}{N-p} \sum_{i=p+1}^{N} [m(\mathbf{X}_i;\boldsymbol{\alpha}) - Y_i]^2.$$

# Convergence

For a wide class of nonlinear systems and aggregative (dictionary) model, it holds that

$$E\{Q(\hat{\boldsymbol{\alpha}})\} - Q(\boldsymbol{\alpha}^*) \leq C\frac{\sqrt{N \ln D}}{N-p}$$

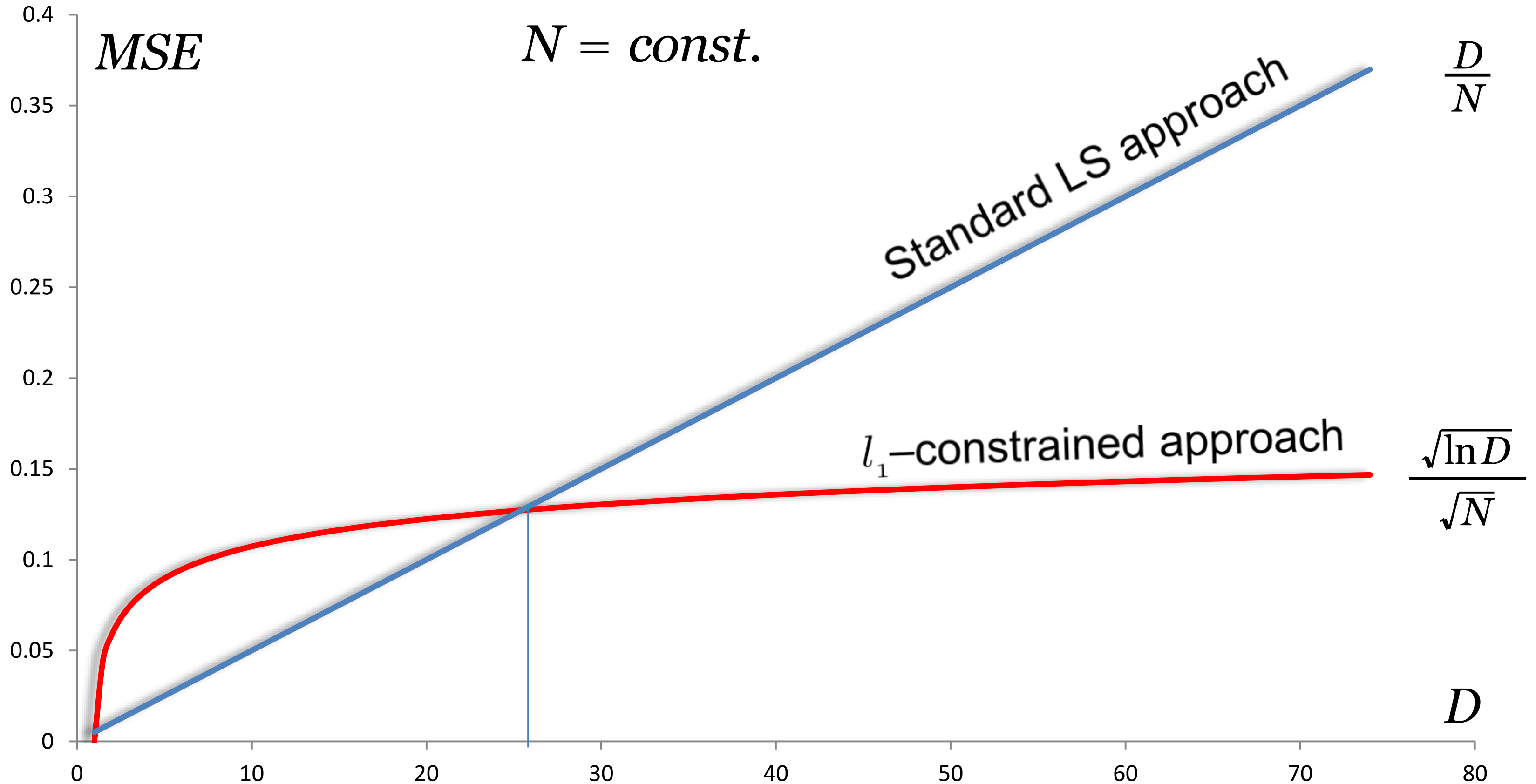
where

$$Q(\boldsymbol{\alpha}) = E\{m(\mathbf{X}_n;\boldsymbol{\alpha}) - Y_n\}^2$$

and where

$$\boldsymbol{\alpha}^* = \operatorname{argmin}_{\boldsymbol{\alpha}\in A} Q(\boldsymbol{\alpha})$$

# $l_1$–constrained vs. standard LS modeling...

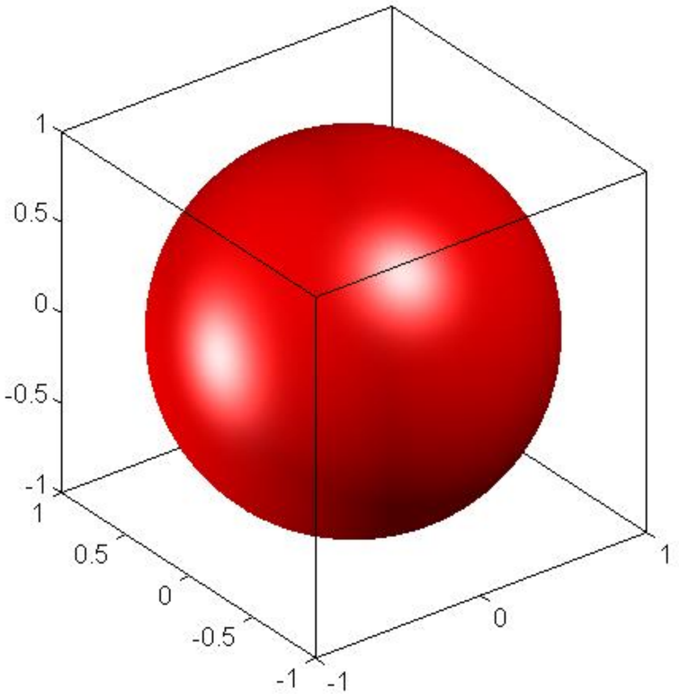



Apply standard approach

Apply $l_1$–constrained approach

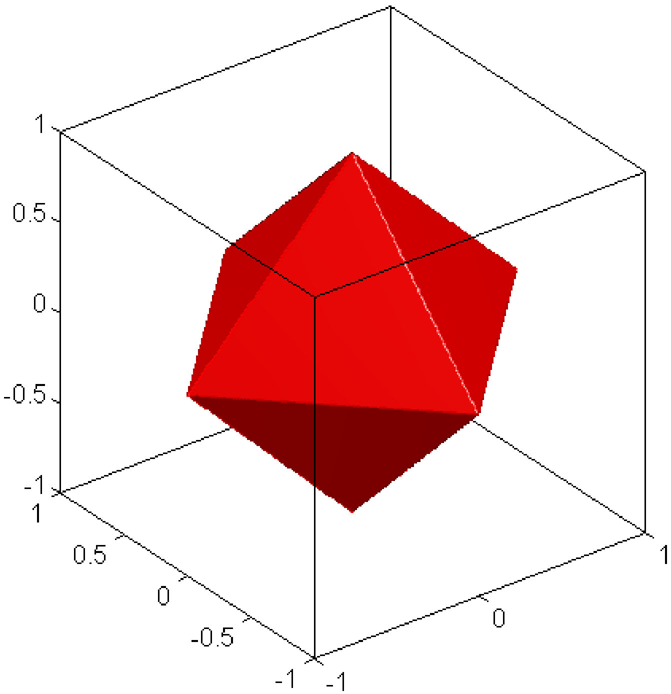
# Why it works?

$l_2$–ball for *D=3*



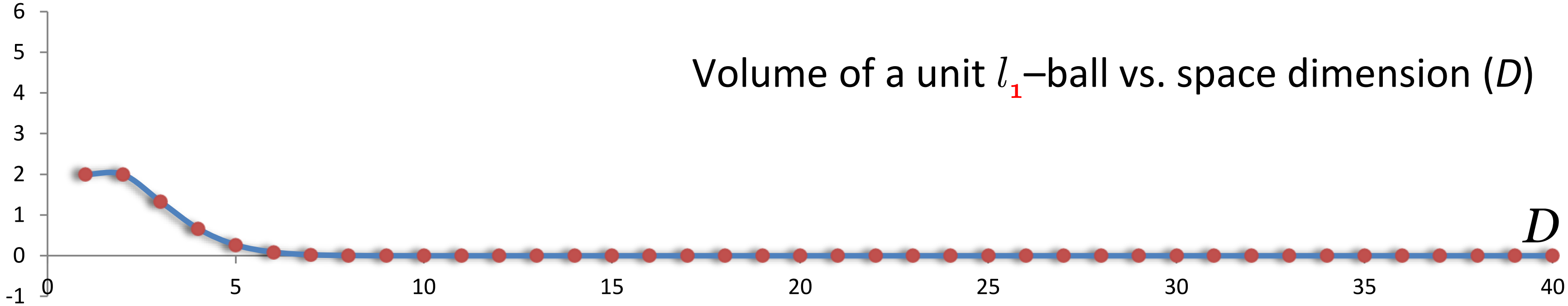

$$V = \frac{4}{3}\pi \approx 4.1888$$

$l_1$–ball for *D=3*

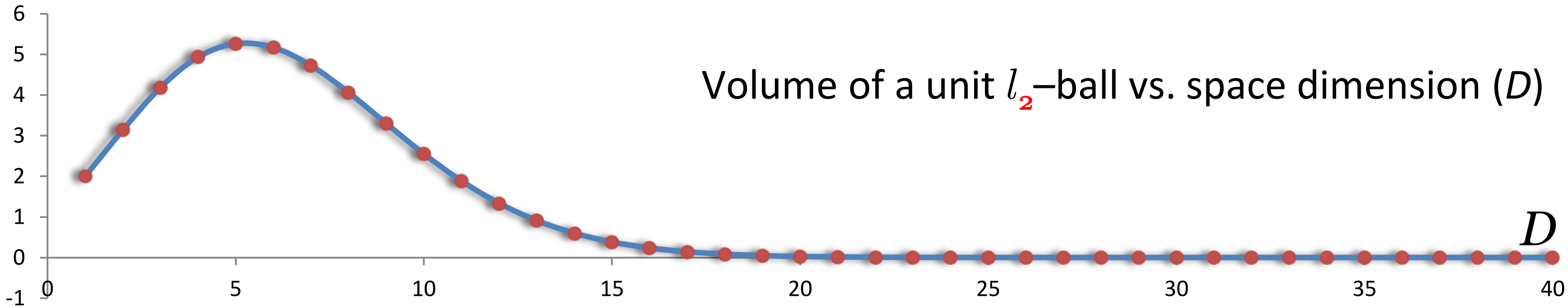



$$V = 1.333$$

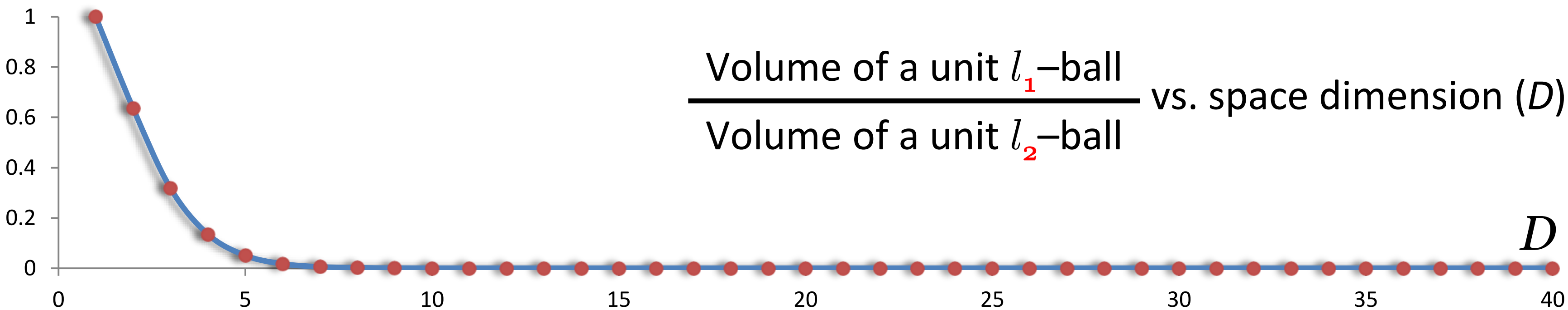
# Why it works?



Volume of a unit $l_1$–ball vs. space dimension ($D$)
$D$
6
5
4
3
2
1
0
-1
0
5
10
15
20
25
30
35
40

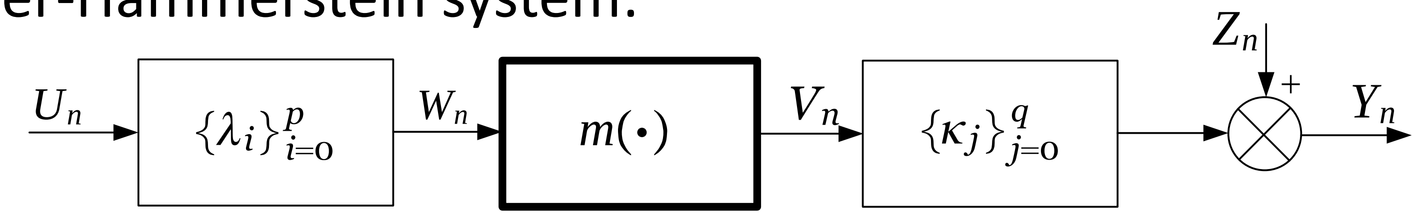
Volume of a unit $l_2$–ball vs. space dimension ($D$)
$D$
6
5
4
3
2
1
0
-1
0
5
10
15
20
25
30
35
40

$\frac{\text{Volume of a unit } l_1\text{–ball}}{\text{Volume of a unit } l_2\text{–ball}}$ vs. space dimension ($D$)
$D$
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
40

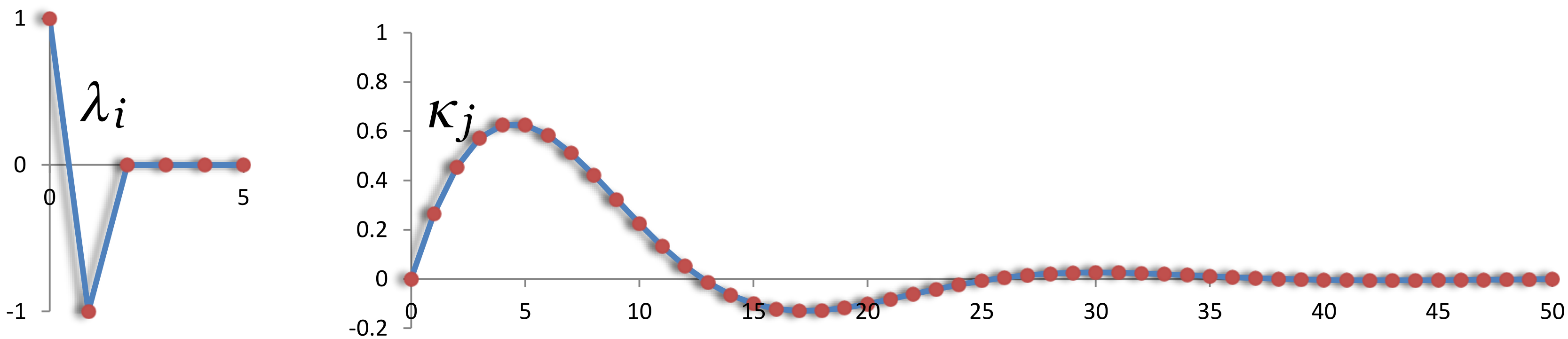
# Experiment settings

Wiener-Hammerstein system:



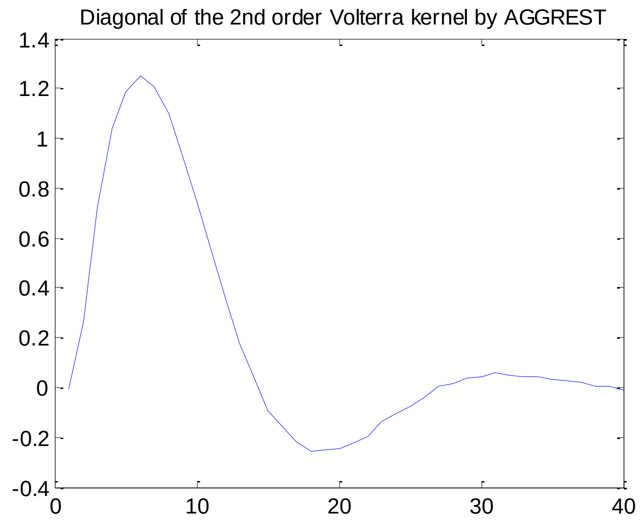

- Nonlinearity: $m(w) = w^2$.

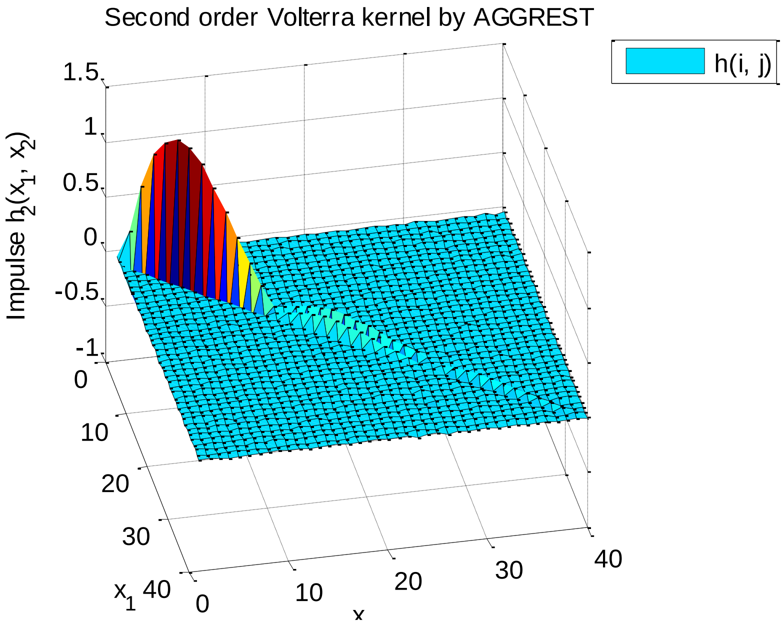
**Impulse responses:**

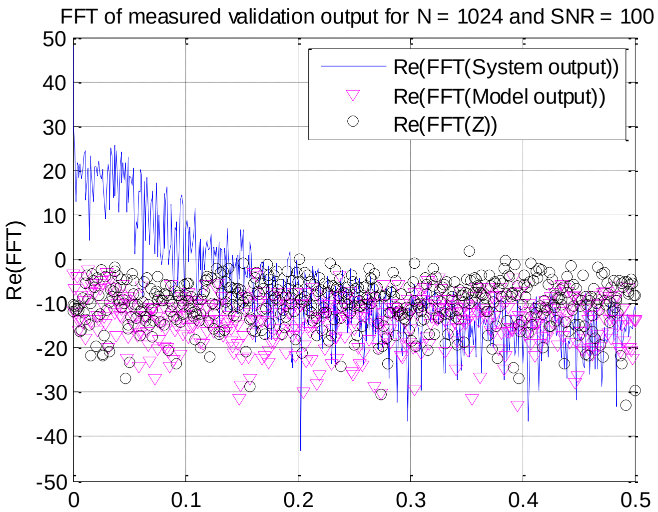


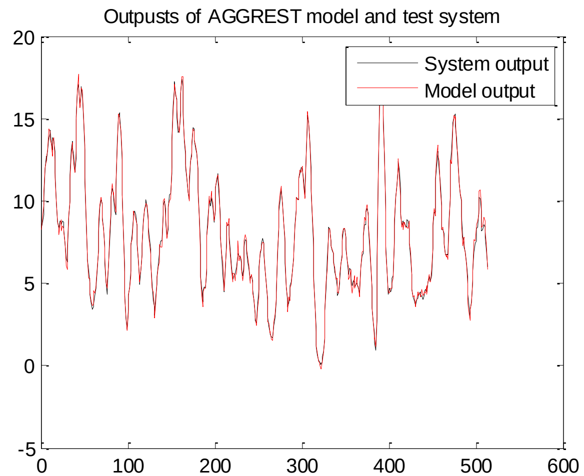

- Input signal: Uniform $(-\sqrt{3}, \sqrt{3})$
- Noise signal: SNR=10
- Dictionary: Unique Volterra series components (L =40, P=2)

# Experiments results



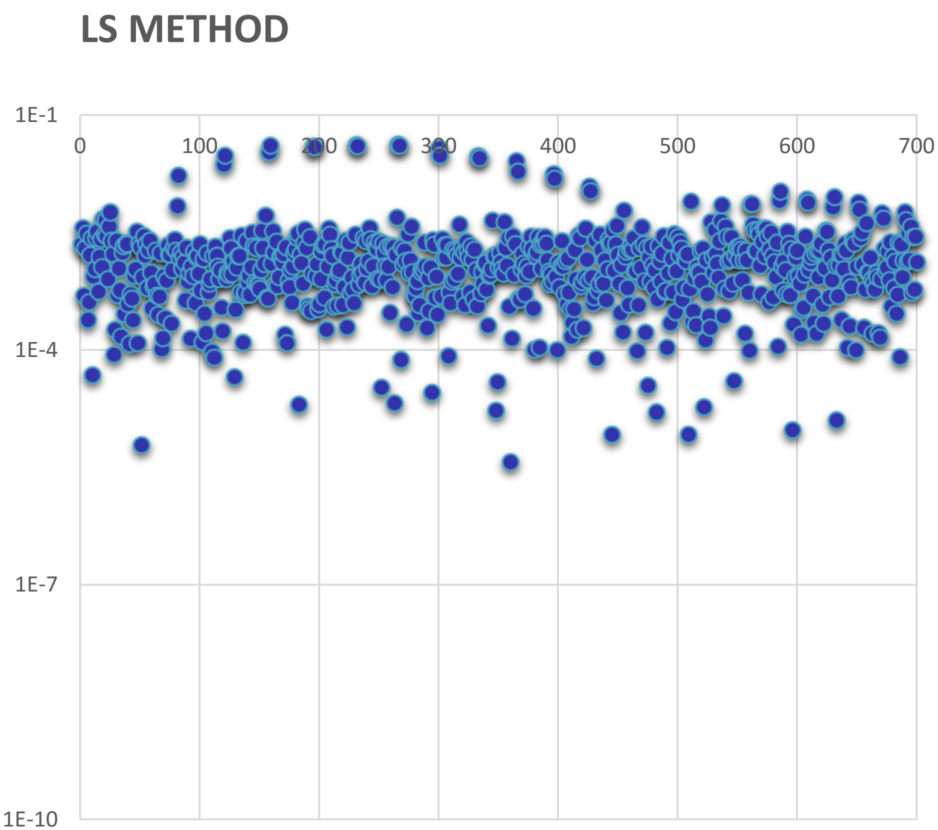



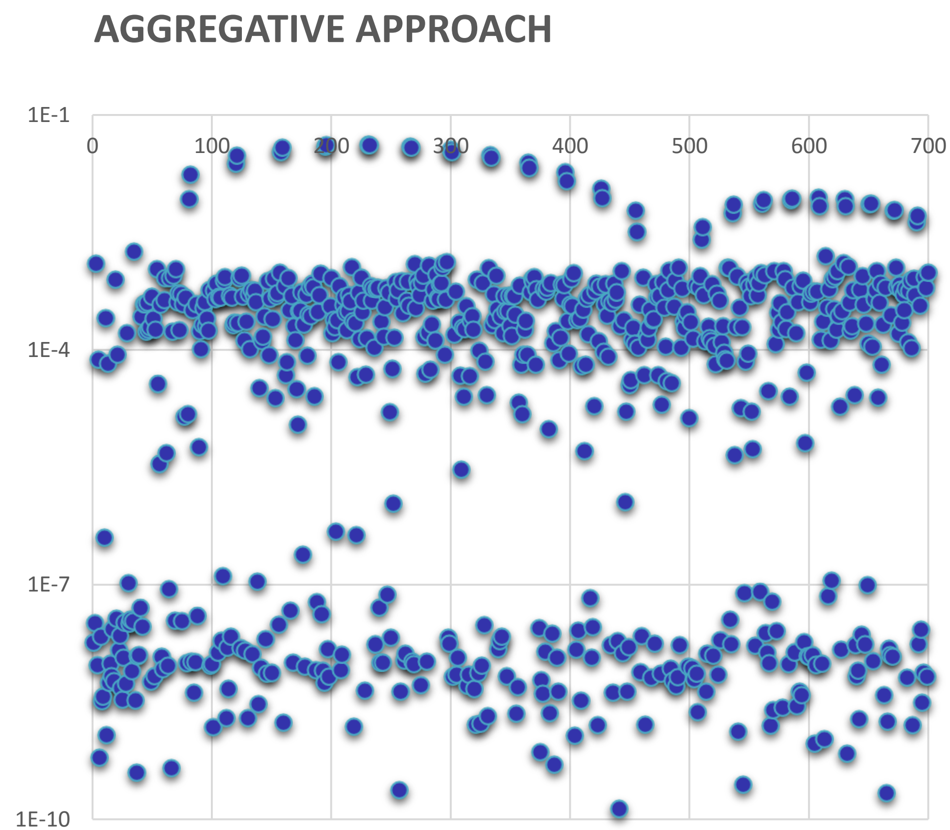



# LS Method vs Aggregative modeling

Dictionary weights evaluated by the LS method and aggregative approach.



# Bibliography


- Greblicki W. and Pawlak M. *Nonparametric system identification*. Cambridge: Cambridge University Press, 2008.
- Greblicki W. and Pawlak M. "Identification of discrete Hammerstein systems using kernel regression estimates." *Automatic Control, IEEE Transactions on* 31.1 (1986): 74-77.
- Greblicki W. "Nonparametric identification of Wiener systems."*Information Theory, IEEE Transactions on* 38.5 (1992): 1487-1493.
- Wachel P. Sliwinski P. and Hasiewicz Z. "Nonparametric identification of MISO Hammerstein system from structured data." *Journal of Systems Science and Systems Engineering* 24.1 (2015): 68-80.
- Hasiewicz Z. and Sliwinski P. "Identification of non-linear characteristics of a class of block-oriented non-linear systems via Daubechies wavelet-based models." *International Journal of Systems Science* 33.14 (2002): 1121-1144.
- Hasiewicz Z. and Mzyk G. "Combined parametric-nonparametric identification of Hammerstein systems." *Automatic Control, IEEE Transactions on* 49.8 (2004): 1370-1375.
- Sliwinski P. *Nonlinear system identification by Haar wavelets*. Vol. 210. Springer Science & Business Media, 2012.
- Mzyk, G. *Combined Parametric-Nonparametric Identification of Block-Oriented Systems*. Berlin: Springer, 2014.
- Wachel P. and Mzyk G. "Direct identification of the linear block in Wiener system." *International Journal of Adaptive Control and Signal Processing* (2015).

Thank you