Model structure learning for LPV system identification using LS-SVMs



Roland Tóth

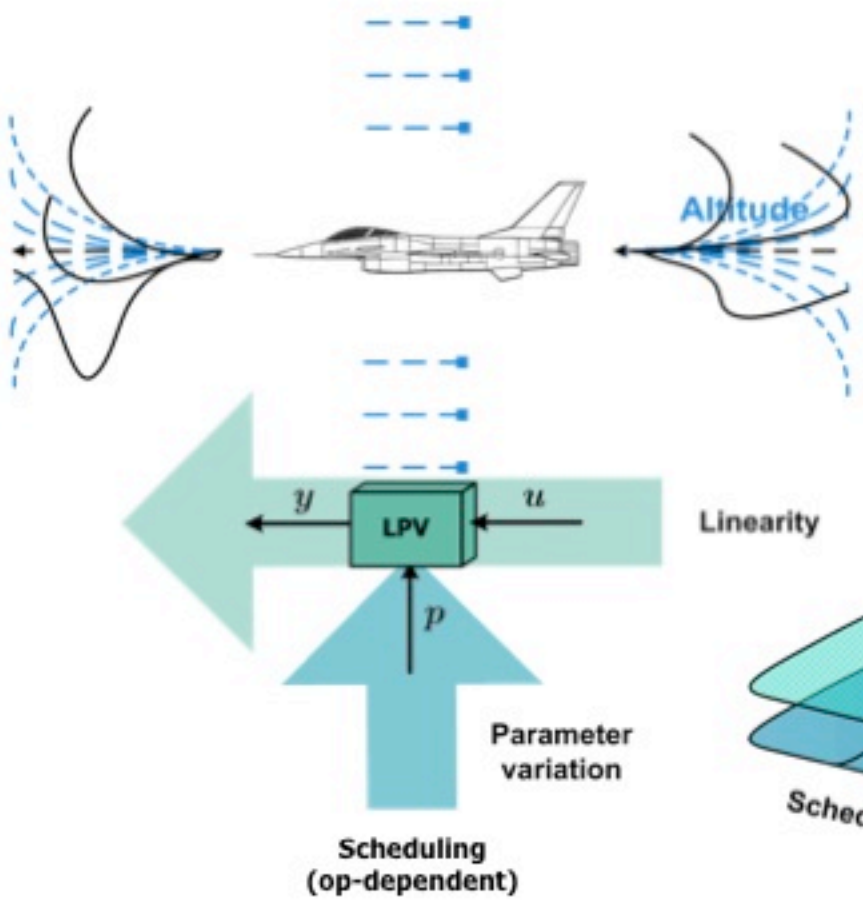
Leuven, July 14, 2014

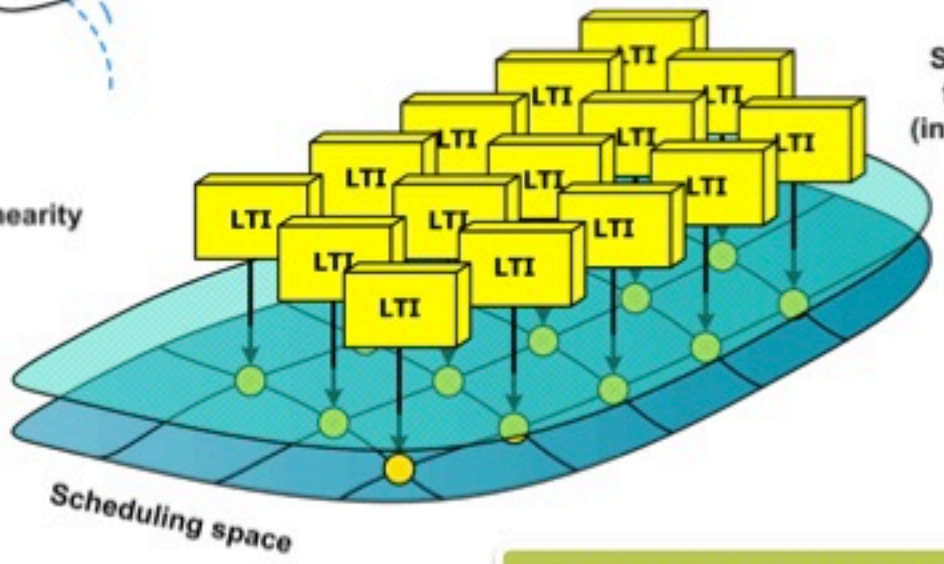
TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

The LPV system concept

Principles and formulation: the local approach





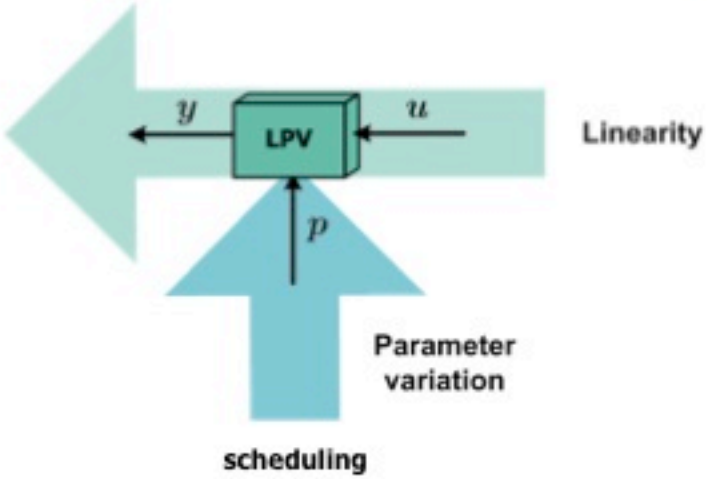
Local modeling principle

Assume p to be exogenous & measurable

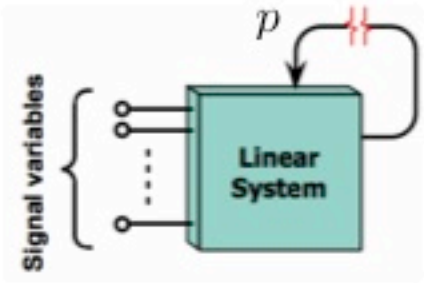
The LPV system concept

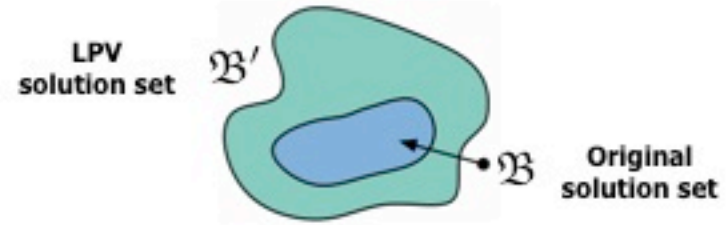
Principles and formulation: the global approach





p becomes "exogenous"





Global modeling principle

The local and global approaches characterize the spectrum of embedding principles.

The LPV system concept

Applications and outlooks








- Many promising applications
  - Flight control and missile autopilots
  - Aeroelsticity and Fly-by-wire systems
  - Robotics
  - Process control
  - Turbofan engines
  - Magnetic bearings & gyro control
  - Automotive systems
  - Energy management (batteries, inverter)
  - Biomechanics
  - Environmental (rain flow, canal models)

Contents

- The LPV system concept
- An LPV lest-squares SVM estimator
- Joint model order selection
- Estimation under general noise conditions
- Conclusions

An LPV lest-squares SVM estimator

The problem of model structure selection

- LPV-ARX model structure (SISO case)

Coefficient functions

Scheduling signal $p : \mathbb{Z} \to \mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$

$$y(k) + \sum_{i=1}^{n_\mathrm{a}} a_i(p(k))y(k-i) = \sum_{j=0}^{n_\mathrm{b}} b_j(p(k))u(k-j) + e(k)$$

Noise (white)

- Linear parameterization:

$$\begin{bmatrix} \phi_1 & \dots & \phi_{n_\mathrm{g}} \end{bmatrix}^\top \triangleq \begin{bmatrix} a_1 & \dots & a_{n_\mathrm{a}} & b_0 & \dots & b_{n_\mathrm{b}} \end{bmatrix}^\top$$

$$n_\mathrm{g} \triangleq n_\mathrm{a} + n_\mathrm{b} + 1$$

An LPV lest-squares SVM estimator

The problem of model structure selection

- LPV-ARX model structure (SISO case)

Coefficient functions

Scheduling signal $p : \mathbb{Z} \to \mathbb{P} \subseteq \mathbb{R}^{n_p}$

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

Noise (white)

- Linear parameterization:

Chosen functions (lin. independent)

$$\phi_i(.) = \theta_{i0} + \sum_{j=1}^{s_i} \theta_{ij}\psi_{ij}(.)$$

Parameters (unknown)

- Can be written as a linear regression:

$$y(k) = \varphi^{\top}(k)\theta + e(k)$$

An LPV lest-squares SVM estimator

The problem of model structure selection

- LPV-ARX model structure (SISO case)

Coefficient functions

Scheduling signal $p : \mathbb{Z} \to \mathbb{P} \subseteq \mathbb{R}^{n_p}$

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

Noise (white)

- The least-square estimate w.r.t. a dataset $\mathcal{D}_N = \{(u(k), p(k), y(k))\}_{k=1}^N$

$$\hat{\theta}_N = \arg\min_{\theta \in \mathbb{R}^{n_\theta}} \frac{1}{N} \sum_{k=1}^{N} (y(k) - \varphi^\top(k)\theta)^2$$

How many parameters are needed for an optimal estimate?

- Can be written as a linear regression:

$$y(k) = \varphi^\top(k)\theta + e(k)$$

Which ψ_{ij} should be used to capture the true system?

An LPV lest-squares SVM estimator

Setting up the regression problem

- The LPV-ARX structure

Problem 1:
How to estimate the coefficients in a "non-parametric" way?

$$y(k) + \sum_{i=1}^{n_\mathrm{a}} a_i(p(k))y(k-i) = \sum_{j=0}^{n_\mathrm{b}} b_j(p(k))u(k-j) + e(k)$$

$$\Downarrow$$

$$y(k) = \sum_{i=1}^{n_\mathrm{g}} \omega_i^\top \phi_i(p_k)x_i(k) + e(k)$$

Parameter vector
$\omega_i \in \mathbb{R}^{n_\mathrm{H}}$

Undefined feature map $\phi_i : \mathbb{R} \to \mathbb{R}^{n_\mathrm{H}}$
(potentially $n_\mathrm{H} = \infty$)

An LPV lest-squares SVM estimator

Setting up the regression problem

- The LPV-ARX structure

Problem 1:
How to estimate the coefficients in a "non-parametric" way?

$$y(k) + \sum_{i=1}^{n_\mathrm{a}} a_i(p(k))y(k-i) = \sum_{j=0}^{n_\mathrm{b}} b_j(p(k))u(k-j) + e(k)$$

$$\Downarrow$$

$$y(k) = \sum_{i=1}^{n_\mathrm{g}} \omega_i^\top \phi_i(p_k)x_i(k) + e(k)$$

- Introduce

$$\begin{aligned} x_i(k) &= y(k-i), \quad i = 1, \ldots, n_\mathrm{a}, \\ x_{n_\mathrm{a}+1+j}(k) &= u(k-j), \quad j = 0, \ldots, n_\mathrm{b} \end{aligned}$$

An LPV lest-squares SVM estimator

Setting up the regression problem

- The LPV-ARX structure

Problem 1:
How to estimate the coefficients in a "non-parametric" way?

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

$$\Downarrow$$

$$y(k) = \omega^\top \varphi(k) + e(k)$$

- Introduce

$$\varphi(k) = \left[\begin{array}{ccc} \phi_1^\top(p_k)x_1(k) & \ldots & \phi_{n_g}^\top(p_k)x_{n_g}(k) \end{array} \right]^\top$$

$$\omega = \left[\begin{array}{cc} \omega_1^\top & \ldots \omega_{n_g}^\top \end{array} \right]^\top \in \mathbb{R}^{n_g n_H}$$

An LPV lest-squares SVM estimator

Ridge regression

- Regularized regression

Reducing function components

Regularization parameter $\gamma \in \mathbb{R}_0^+$

$$\min_{\omega,e} \quad \mathcal{J}(\omega, e) = \frac{1}{2}\sum_{i=1}^{n_g} \omega_i^\top \omega_i + \frac{\gamma}{2}\sum_{k=1}^{N} e^2(k)$$

$$\text{s.t.} \quad e(k) = y(k) - \sum_{i=1}^{n_g} \omega_i^\top \phi_i(p(k)) x_i(k)$$

- Primal solution

$$\hat{\omega} = \underbrace{\left(\Phi^\top \Phi + \gamma^{-1} I\right)}_{\in \mathbb{R}^{n_g n_H \times n_g n_H}}{}^{-1} \Phi^\top Y$$

The primal solution is not useful!

We want to estimate both ω and $\{\phi_i\}_{i=1}^{n_g}$.

$$Y = \begin{bmatrix} y(1) & y(2) & \dots & y(N) \end{bmatrix}^\top \in \mathbb{R}^N$$

$$\Phi = \begin{bmatrix} \varphi(1) & \varphi(2) & \dots & \varphi(N) \end{bmatrix}^\top \in \mathbb{R}^{N \times n_g n_H}$$

An LPV lest-squares SVM estimator

Dual solution

- Regularized regression (Langrangian form)

$$\mathcal{L}(\omega, e, \alpha) = \mathcal{J}(\omega, e) - \sum_{k=1}^{N} \alpha_k \left(\sum_{i=1}^{n_g} \omega_i^\top \phi_i(k) x_i(k) + e(k) - y(k) \right)$$

- Dual form (based on the KKT conditions):

$$y(k) = \sum_{i=1}^{n_g} \underbrace{\left(\sum_{k=1}^{N} \alpha_k x_i(k) \phi_i^\top(k) \right)}_{\omega_i^\top} \phi_i(k) x_i(k) + \underbrace{\gamma^{-1} \alpha_k}_{e(k)}$$

Inner products:

$$[\Omega^i]_{j,k} = x_i(j) \phi_i^\top(j) \phi_i(k) x_i(k)$$

$$[\Omega]_{j,k} = \sum_{i=1}^{n_g} [\Omega^i]_{j,k}$$

Parameter:

$$\alpha = [\alpha_1 \ \dots \ \alpha_N]^\top \in \mathbb{R}^N$$

An LPV lest-squares SVM estimator

Dual solution

- Regularized regression (Langrangian form)

$$\mathcal{L}(\omega, e, \alpha) = \mathcal{J}(\omega, e) - \sum_{k=1}^{N} \alpha_k \left(\sum_{i=1}^{n_g} \omega_i^\top \phi_i(k) x_i(k) + e(k) - y(k) \right)$$

- Dual form (based on the KKT conditions):

$$Y = \underbrace{\left(\Omega + \gamma^{-1} I\right)}_{\in \mathbb{R}^{N \times N}} \alpha$$

Inner products:

$$[\Omega^i]_{j,k} = x_i(j) \phi_i^\top(j) \phi_i(k) x_i(k)$$

$$[\Omega]_{j,k} = \sum_{i=1}^{n_g} [\Omega^i]_{j,k}$$

Parameter:

$$\alpha = [\alpha_1 \ \ldots \ \alpha_N]^\top \in \mathbb{R}^N$$

An LPV lest-squares SVM estimator

The kernel trick

- The LS-SVM in the LPV case:

$$Y = \underbrace{\left(\Omega + \gamma^{-1} I\right)}_{\in \mathbb{R}^{N \times N}} \alpha$$

$$\begin{aligned}
[\Omega^i]_{j,k} &= x_i(j)\phi_i^\top(j)\phi_i(k)x_i(k) \\
&= x_i(j)\langle \phi_i(j), \phi_i(k)\rangle x_i(k) \\
&= x_i(j)\big(K_i(p(j), p(k))\big)x_i(k)
\end{aligned}$$

Instead of defining $\{\phi_i^\top(\cdot)\}_{i=1}^{n_g}$ → Define the inner products $\phi_i^\top(j)\phi_i(k)$ → Positive definite kernel functions $K_i(\cdot,\cdot)$

An LPV lest-squares SVM estimator

The kernel trick

- Radial Basis Function (RBF) kernel

$$K_i(p(j), p(k)) = \exp\left(-\frac{\|p(j) - p(k)\|_2^2}{\sigma_i^2}\right)$$

other kernels: polynomial, Direchlet, etc. can also be used.

- Solution

$$\hat{\alpha} = \left(\Omega + \gamma^{-1} I\right)^{-1} Y$$

- Function estimates:

$$\hat{j} = j + n_{\mathrm{a}} + 1$$

$$\hat{a}_i(.) = \omega_i^\top \phi_i(.) = \sum_{k=1}^{N} \hat{\alpha}_k K_i(p(k), .) x_i(k)$$

$$\hat{b}_j(.) = \omega_{\hat{j}}^\top \phi_{\hat{j}}(.) = \sum_{k=1}^{N} \hat{\alpha}_k K_{\hat{j}}(p(k), .) x_{\hat{j}}(k)$$

An LPV lest-squares SVM estimator

Estimation properties

- Recovery of coefficient dependencies:

$$a_i \in \Pi_i = \mathrm{Span}_{\mathbf{p}\in\mathbb{P}}\{K_i(\mathbf{p},.)\}$$

Proper selection (optimization) of kernels

- Stochastic properties (bias):

Biased if $e(k)$ is non-white!

$$\bar{\mathbb{E}}\left\{\hat{\phi}_i(.)-\phi_{i,\mathrm{o}}(.)\right\} = \underbrace{\bar{\mathbb{E}}\left\{\frac{\gamma}{N}K_{.x}^{(i)}E\right\}-\bar{\mathbb{E}}\left\{\frac{\gamma}{N}K_{.x}^{(i)}\left(\frac{1}{N}K_{xx}^{(i)}+\gamma^{-1}I_N\right)^{-1}K_{xx}^{(i)}E\right\}}_{\text{noise bias}}$$

$$+\underbrace{\bar{\mathbb{E}}\left\{\frac{1}{\gamma}\left(\frac{1}{N}K_{xx}^{(i)}+\gamma^{-1}I_N\right)^{-1}\alpha_{\mathrm{o}}\right\}}_{\text{regularization bias}}$$

- Stochastic properties (variance):

Governed by the LS properties

An LPV lest-squares SVM estimator

Example: LPV-ARX identification

- Consider the LPV data-generating system

$$y(k) = -a_1(p_k)y(k-1) + \sum_{i=0}^{1} b_i(p_k)u(k-i) + e_o(k)$$

Noise (white, Gaussian)

with $\mathbb{P} = [-1, 1]$ and

$$a_1(p_k) = 0.1 \cdot \mathrm{sinc}(\pi^2 p_k)$$

$$b_0(p_k) = \begin{cases} +0.5 & \text{if } p_k > +0.5 \\ p_k & \text{if } |p_k| \leq 0.5 \\ -0.5 & \text{if } p_k < -0.5 \end{cases}$$

$$b_1(p_k) = -0.2 \cdot p_k^2$$

Estimate the nonlinear dependencies!

An LPV lest-squares SVM estimator

Example: LPV-ARX identification

- Data records: white input, scheduling and noise signals with distributions $u(k) \in \mathcal{U}(-1,1)$, $p(k) \in \mathcal{U}(-1,1)$ and SNR=10dB.
- 100 Monte-Carlo runs:
  - Optimized RBF kernels K_1, K_2, K_3 with

$$\sigma_1 = \sigma_2 = \sigma_3 = 0.7$$

  - Tuned regularization parameter

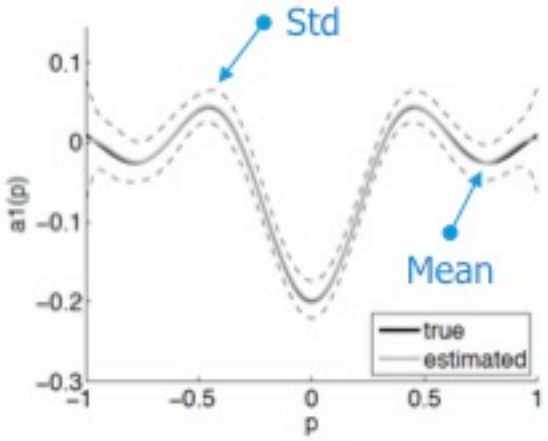
$$\gamma = 10^4$$

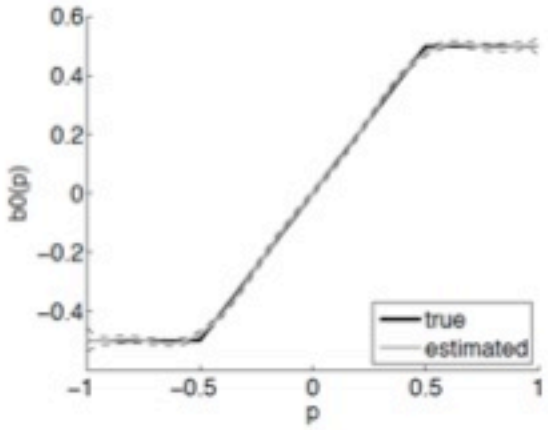
  - 100 data records $\mathcal{D}_N$ with $N = 1500$ samples
  - Dynamic structure (model order) is assumed to be known

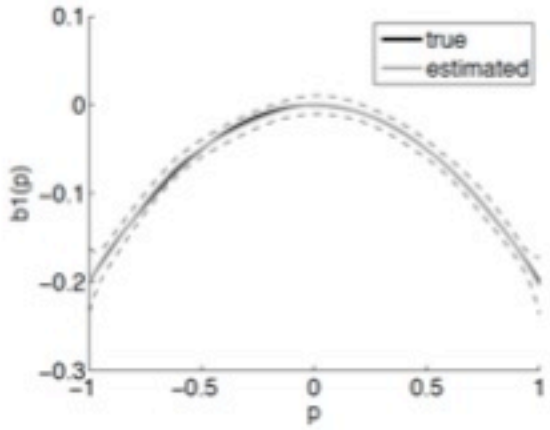
An LPV lest-squares SVM estimator

Example: LPV-ARX identification

- Function estimates







- Fit score of the simulated (noise-free) output:

$$\mathrm{BFR} := 100\% \cdot \max\left(1 - \frac{\|y - \hat{y}\|_2}{\|y - \bar{y}\|_2}, 0\right)$$

	Mean	STD
LS-SVM	95.22	0.005

Contents

- The LPV system concept
- An LPV lest-squares SVM estimator
- Joint model order selection
- Estimation under general noise conditions
- Conclusions

Joint model order selection

Including an order selection objective

- The LPV-ARX structure

Coefficient function estimation is solved!

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

How to estimate the model order in this "non-parametric" setting?

Idea:
Apply additional regularization on the number of estimated functions.
(Multiple Kernel Learning with sparsity)

Joint model order selection

Concept of sparsity (primal form)

- What do we want with regularization?
  - Estimate the support of θ_o (recover sparsity):

$$\underset{\theta \in \mathbb{R}^{n_\theta}}{\text{minimize}} \quad \|\theta\|_0$$

Assumption:

$$\mathcal{S}_o \in \mathcal{M}$$

Non-convex

Joint model order selection

Concept of sparsity (primal form)

- What do we want with regularization?
  - Estimate the support of θ_o (recover sparsity):

$$\underset{\theta \in \mathbb{R}^{n_\theta}}{\text{minimize}} \quad \|\theta\|_1$$

  - Minimize the prediction error:

$$\min_{\theta \in \mathbb{R}^{n_\theta}} V_N(\theta, \mathcal{D}_N) = \frac{1}{N} \|y(k) - \varphi^\top(k)\theta\|_{\ell_2}^2$$

  - Sparse optimization problem

$$\begin{aligned} \min_{\theta \in \mathbb{R}^{n_\theta}} \quad & \|\theta\|_1 \\ \text{s.t.} \quad & V_N(\theta, \mathcal{D}_N) < \varepsilon_* \end{aligned}$$

Assumption:
$\mathcal{S}_o \in \mathcal{M}$

Convex

Contradicting objectives

A particularly interesting variant: SPARSEVA

Joint model order selection

Including an order selection objective

- The LPV-ARX structure

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

How to estimate the model order in this "non-parametric" setting?

Minimize an ℓ_1 objective

Introduce a gridding:

$$\mathcal{P} = \{m_s\}_{s=1}^{M} \subset \mathbb{P}$$

"Coef. measure:"

$$\overline{\eta}_i = \max_{s=1,\ldots,M} \left|\omega_i^{\top} \phi_i(m_s)\right|$$

Joint model order selection

Including an order selection objective

- The LPV-ARX structure

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

How to estimate the model order in this "non-parametric" setting?

- Apply an elastic net type of cost function:

$$\mathcal{J}(\omega, e) = \frac{1}{2}\sum_{i=1}^{n_g} \omega_i^\top \omega_i + \frac{\gamma}{2}\sum_{k=1}^{N} e^2(k) + \mu \left\| \left[\overline{\eta}_1 \ \ldots \ \overline{\eta}_{n_g} \right]^\top \right\|_1$$

Joint model order selection

The optimization problem

- The resulting QP:

$$\begin{aligned}
\min_{\omega,e,r} \quad & \mathcal{J}(\omega,e) = \frac{1}{2}\sum_{i=1}^{n_g} \omega_i^\top \omega_i + \frac{\gamma}{2}\sum_{k=1}^{N} e^2(k) + \mu \sum_{i=1}^{n_g} r_i \\
\text{s.t.} \quad & e(k) = y(k) - \sum_{i=1}^{n_g} \omega_i^\top \phi_i(p(k)) x_i(k) \\
& -r_i \leq \omega_i^\top \phi_i(m_s) \leq r_i, \ \forall s \in \mathbb{I}_1^M
\end{aligned}$$

Joint model order selection

The optimization problem

- Via KKT, the Lagrangian dual is

$$g(\alpha, \beta^+, \beta^-) = -\frac{1}{2\gamma}\alpha^\top\alpha + \alpha^\top Y - \frac{1}{2}\sum_{i=1}^{n_g} \begin{bmatrix} \alpha \\ \beta_i^- - \beta_i^+ \end{bmatrix}^\top \Omega_i \begin{bmatrix} \alpha \\ \beta_i^- - \beta_i^+ \end{bmatrix}$$

- Under the (dual) constraints

$$\mu = \sum_{s=1}^{M} (\beta_{i,s}^- + \beta_{i,s}^+)$$

$$0 \leq \beta_{i,s}^-, \beta_{i,s}^+$$

$$\beta_i^- = \begin{bmatrix} \beta_{i,1}^- & \cdots & \beta_{i,M}^- \end{bmatrix}^\top$$

$$\beta_i^+ = \begin{bmatrix} \beta_{i,1}^+ & \cdots & \beta_{i,M}^+ \end{bmatrix}^\top$$

Joint model order selection

The optimization problem

- Via KKT, the Lagrangian dual is

$$g(\alpha, \beta^+, \beta^-) = -\frac{1}{2\gamma}\alpha^\top\alpha + \alpha^\top Y - \frac{1}{2}\sum_{i=1}^{n_g} \begin{bmatrix} \alpha \\ \beta_i^- - \beta_i^+ \end{bmatrix}^\top \Omega_i \begin{bmatrix} \alpha \\ \beta_i^- - \beta_i^+ \end{bmatrix}$$

- Under the (dual) constraints

$$\mu = \sum_{s=1}^{M} (\beta_{i,s}^- + \beta_{i,s}^+)$$
$$0 \leq \beta_{i,s}^-, \beta_{i,s}^+$$

$$X_i = \mathrm{diag}(x_i(1), \ldots x_i(N))$$
$$\Phi_i^{(N)} = \begin{bmatrix} \phi_i(p(1)) & \cdots & \phi_i(p(N)) \end{bmatrix}$$
$$\Phi_i^{(M)} = \begin{bmatrix} \phi_i(m_1) & \cdots & \phi_i(m_M) \end{bmatrix}$$

Joint model order selection

The optimization problem

- Via KKT, the Lagrangian dual is

$$g(\alpha, \beta^+, \beta^-) = -\frac{1}{2\gamma}\alpha^\top\alpha + \alpha^\top Y - \frac{1}{2}\sum_{i=1}^{n_g} \begin{bmatrix} \alpha \\ \beta_i^- - \beta_i^+ \end{bmatrix}^\top \Omega_i \begin{bmatrix} \alpha \\ \beta_i^- - \beta_i^+ \end{bmatrix}$$

- Under the (dual) constraints

$$\mu = \sum_{s=1}^{M} (\beta_{i,s}^- + \beta_{i,s}^+)$$
$$0 \leq \beta_{i,s}^-, \beta_{i,s}^+$$

$$\Omega_i = \begin{bmatrix} X_i(\Phi_i^{(N)})^\top \\ (\Phi_i^{(M)})^\top \end{bmatrix} \begin{bmatrix} \Phi_i^{(N)} X_i \Phi_i^{(M)} \end{bmatrix}$$

Joint model order selection

The resulting RLS-SVM:

- The dual QP:

$$\begin{aligned} \min_{\alpha,\beta^+,\beta^-} \quad & -g(\alpha,\beta^+,\beta^-) \\ \text{s.t.} \quad & \mu = \sum_{s=1}^{M} \left(\beta_{i,s}^- + \beta_{i,s}^+\right) \\ & 0 \leq \beta_{i,s}^-, \beta_{i,s}^+ \end{aligned}$$

Lesson from sparse parametric ID: Thresholding & re-estimation!

- Resulting estimates:

$$\hat{a}_i(\cdot) = \omega_i^\top \phi_i(\cdot) = \sum_{k=1}^{N} \alpha_k K_i(p(k),\cdot)x_i(k) + \sum_{s=1}^{M} \left(\beta_{i,s}^- - \beta_{i,s}^+\right) K_i(m_s,\cdot),$$

$$\hat{b}_j(\cdot) = \omega_{\hat{j}}^\top \phi_{\hat{j}}(\cdot) = \sum_{k=1}^{N} \alpha_k K_{\hat{j}}(p(k),\cdot)x_{\hat{j}}(k) + \sum_{s=1}^{M} \left(\beta_{\hat{j},s}^- - \beta_{\hat{j},s}^+\right) K_{\hat{j}}(m_s,\cdot),$$

$$i \in \mathbb{I}_1^{n_a}, \quad \hat{j} = j + n_a + 1, j \in \mathbb{I}_0^{n_b}$$

Joint model order selection

Example: LPV-ARX identification

- Consider the LPV data-generating system

$$y(k) = a_1(p_k)y(k-1) + a_2(p_k)y(k-2) + b_5(p_k)u(k-5) + e_o(k)$$

Noise (white, Gaussian)

with $\mathbb{P} = [-1, 1]$ and

$$a_1(p_k) = \begin{cases} -0.5, & \text{if } p_k > 0.5; \\ -p_k, & \text{if } -0.5 \leq p_k \leq 0.5; \\ 0.5, & \text{if } p_k < -0.5; \end{cases}$$

$$a_2(p_k) = \sin(2\pi p_k)$$

$$b_5(p_k) = p_k^3$$

Let's consider now order selection!

An LPV lest-squares SVM estimator

Example: LPV-ARX identification

- Data records: white input, scheduling and noise signals with distributions $u(k) \in \mathcal{U}(-1,1)$, $p(k) \in \mathcal{U}(-1,1)$ and SNR=12dB.
- 100 Monte-Carlo runs:
  - Optimized RBF kernels K_1, K_2, K_3 for the LS-SVM, resulting in

    $$\sigma_1 = \sigma_2 = \sigma_3 = 0.6$$

  - Tuned regularization parameters

    $$\lambda = 600, \mu = 350$$

  - 100 data records $\mathcal{D}_N$ with $N = 500$ samples
  - Gridding:

    $$\mathcal{P} = \{m_s\}_{s=1}^{7} = \{-1,\ -0.66,\ -0.33,\ 0,\ 0.33,\ 0.66,\ 1\}$$

Assumed order:

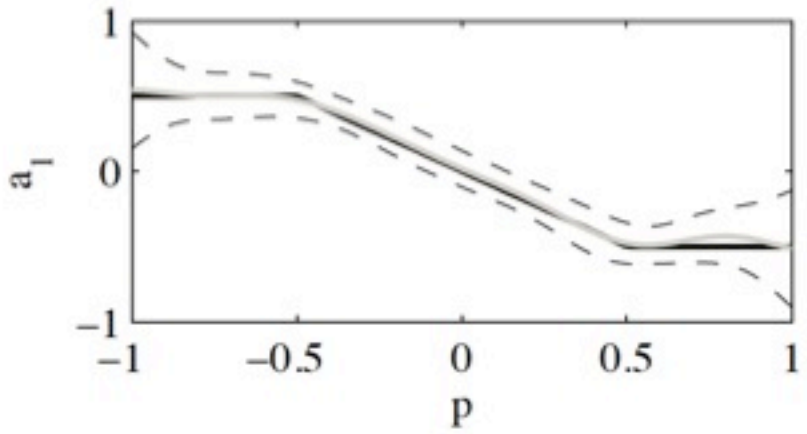
$$n_a = n_b = 10$$

An LPV lest-squares SVM estimator

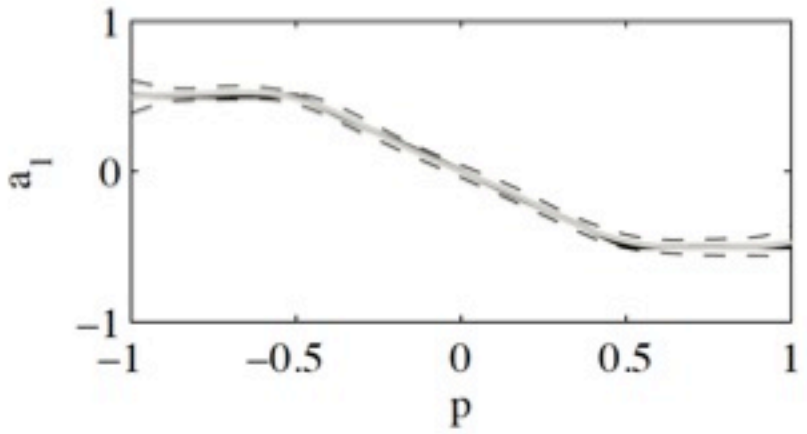
Example: LPV-ARX identification

- Function estimates

	$\bar{a}_1$	$\bar{a}_2$	$\bar{a}_3$	$\bar{a}_4$	$\bar{a}_5$	$\bar{a}_6$	$\bar{a}_7$	$\bar{a}_8$	$\bar{a}_9$	$\bar{a}_{10}$
True value	0.5	1	0	0	0	0	0	0	0	0
Mean (LS-SVM)	$8.3\cdot10^{-1}$	1.1	$4.4\cdot10^{-1}$	$4.4\cdot10^{-1}$	$4.8\cdot10^{-1}$	$4.6\cdot10^{-1}$	$4.7\cdot10^{-1}$	$4.6\cdot10^{-1}$	$4.8\cdot10^{-1}$	$3.9\cdot10^{-1}$
Mean (RLS-SVM)	$3.2\cdot10^{-1}$	$6.3\cdot10^{-1}$	$2.7\cdot10^{-4}$	$3.8\cdot10^{-4}$	$1.1\cdot10^{-4}$	$9.3\cdot10^{-5}$	$3.4\cdot10^{-5}$	$7.4\cdot10^{-4}$	$3.1\cdot10^{-4}$	$6.2\cdot10^{-5}$
std (LS-SVM)	$2.1\cdot10^{-1}$	$1.0\cdot10^{-1}$	$2.1\cdot10^{-1}$	$2.1\cdot10^{-1}$	$2.5\cdot10^{-1}$	$2.2\cdot10^{-1}$	$2.3\cdot10^{-1}$	$2.1\cdot10^{-1}$	$2.1\cdot10^{-1}$	$1.9\cdot10^{-1}$
std (RLS-SVM)	$1.2\cdot10^{-1}$	$1.4\cdot10^{-1}$	$8.6\cdot10^{-5}$	$1.1\cdot10^{-4}$	$5.2\cdot10^{-5}$	$3.8\cdot10^{-5}$	$7.3\cdot10^{-6}$	$2.1\cdot10^{-4}$	$1.0\cdot10^{-4}$	$1.1\cdot10^{-5}$



(a) LS-SVM



(b) RLS-SVM

Joint model order selection

Recap on the original problem

- The LPV-ARX structure

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + e(k)$$

Can we also constrain the complexity of the estimated functions?

- Introduce coefficient sensitivity:

$$\mathcal{W}(\omega, e) = \frac{1}{2}\|\omega\|_{\ell_2}^2 + \frac{\gamma_1}{2}\|e(k)\|_{\ell_2}^2 + \gamma_2 \sum_{i=1}^{n_g}\sum_{j=1}^{n_p} \left\| \frac{\partial \omega_i \phi_i(p)}{\partial p_j} \right\|_{l_\infty}$$

However, this is an other story…

Contents

- The LPV system concept
- An LPV lest-squares SVM estimator
- Joint model order selection
- Estimation under general noise conditions
- Conclusions

Estimation under general noise conditions

How to get rid of the bias?

- General noise setting

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + v_o(k)$$

Noise (colored)



Correlation between $y(k-i)$ and $v_o(k)$ will introduce bias (rotation) of the estimates!

Previous work:

J. Suykens and J. Vandewalle, "Recurrent least squares support vector machines," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 47, no. 7, pp. 1109–1114, 2000.

T. Falck, J. Suykens, and B. De Moor, "Linear parametric noise models for least squares support vector machines," in Proc. of the 49th IEEE Conf. on Decision and Control, Atlanta, USA, Dec. 2010, pp. 6389–6394.

Estimation under general noise conditions

How to get rid of the bias?

- General noise setting

$$y(k) + \sum_{i=1}^{n_a} a_i(p(k))y(k-i) = \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + v_o(k)$$

Noise (colored)



Correlation between $y(k-i)$ and $v_o(k)$ will introduce bias (rotation) of the estimates!

Idea:
Introduce an instrumental variable to remove the correlation effect.

$$\mathbb{E}\{\zeta(k)v_o(k)\} = 0, \quad \forall k \in \mathbb{Z}$$

Instrument (user defined)

Estimation under general noise conditions

Formulation of the instrumental variable SVM

- Instrumental Ridge regression

Instruments

$$\mathcal{W}(\theta, e) = \frac{1}{2}\theta^\top\theta + \frac{\gamma}{2N^2}\| \sum_{k=1}^{N} \zeta(k)e(k)\|_{\ell_2}^2$$

$$= \frac{1}{2}\|\theta\|_{\ell_2}^2 + \frac{\gamma}{2N^2}\|\Gamma^\top E\|_{\ell_2}^2,$$

where

$$\zeta(k) = \left[\begin{array}{cccc} 1 & \phi_1^\top\left(\xi_1(k)\right) & \dots & \phi_{n_g}^\top\left(\xi_{n_g}(k)\right) \end{array} \right]^\top$$

Generator instruments

This choice allows to apply the same Kernel trick

Estimation under general noise conditions

Formulation of the instrumental variable SVM

- Langrangian:

$$\mathcal{L}(\theta, e, \alpha) = \mathcal{W}(\theta, e) - \sum_{k=1}^{N} \alpha_k \left(\varphi^{\top}(k)\theta + e(k) - y(k)\right)$$

- Dual form (based on the KKT conditions):

Inner product

Inner product

$$\alpha = \frac{\gamma}{N^2} \Gamma \Gamma^{\top} \left(Y - \Phi \Phi^{\top} \alpha\right)$$

$$\Gamma = \left[\begin{array}{ccc} \zeta(1) & \dots & \zeta(N) \end{array} \right]^{\top}$$

Estimation under general noise conditions

Formulation of the instrumental variable SVM

- Langrangian:

$$\mathcal{L}(\theta, e, \alpha) = \mathcal{W}(\theta, e) - \sum_{k=1}^{N} \alpha_k \left(\varphi^{\top}(k)\theta + e(k) - y(k) \right)$$

- Dual form (based on the KKT conditions):

Kernel (instrumental variables)

$$\alpha = \left(\frac{1}{N^2} HG + \gamma^{-1} I_N \right)^{-1} \frac{1}{N^2} HY$$

Kernel (signal variables)

- Estimates:

$$\hat{a}_i(.) = \omega_i^{\top} \phi_i(.) = \sum_{k=1}^{N} \hat{\alpha}_k K_i(p(k), .) x_i(k)$$

$$\hat{b}_j(.) = \omega_j^{\top} \phi_j(.) = \sum_{k=1}^{N} \hat{\alpha}_k K_j(p(k), .) x_j(k)$$

Estimation under general noise conditions

Formulation of the instrumental variable SVM

- Choice of the instrument
  - Should be not correlated with the noise
  - Should be maximally correlated with the process (noise-free) dynamics

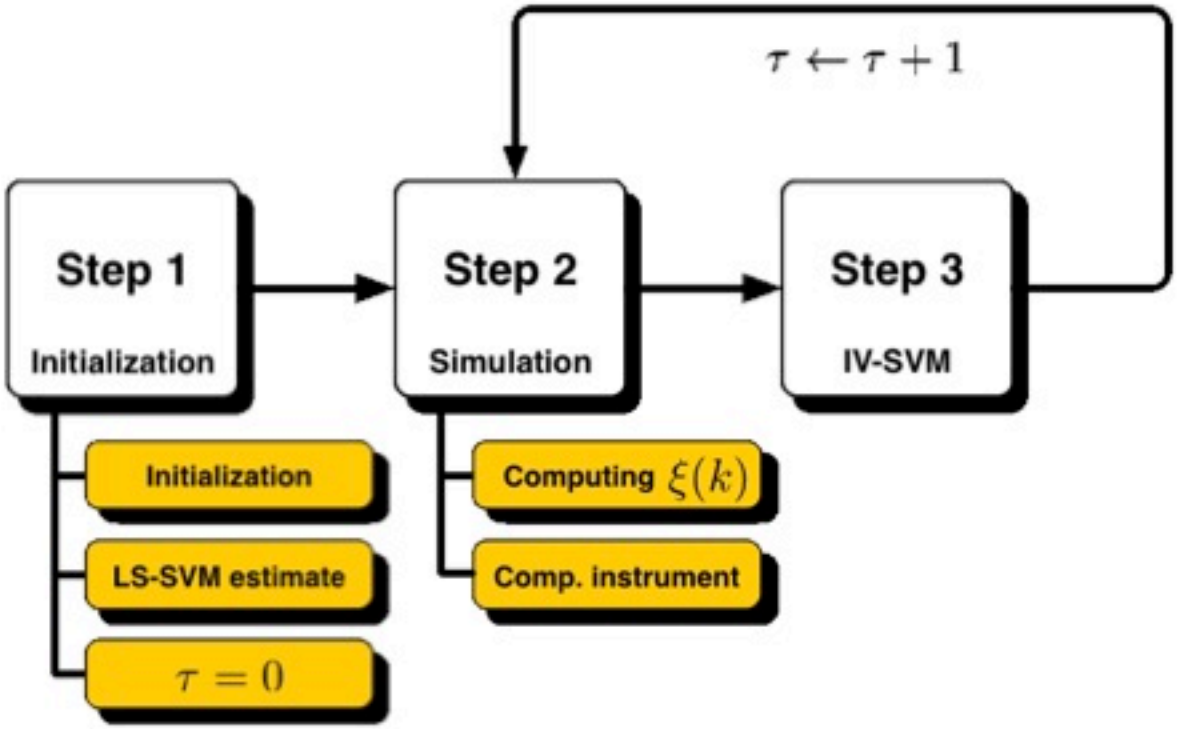
Noise free output signal (unknown)

$$\xi_i(k) = \breve{y}(k-i), \quad i = 1, \ldots, n_\mathrm{a}$$
$$\xi_{n_\mathrm{a}+1+j}(k) = u(k-j), \quad j = 0, \ldots, n_\mathrm{b}$$

Estimation under general noise conditions

The IV-SVM

- The algorithm (refined instrumental variable SVM)



Estimation under general noise conditions

Example: LPV-BJ identification

- Consider the LPV data-generating system

$$y(k) = -0.7y(k-1) + a_2(p_k)y(k-2) + b_0(p_k)u(k) - 0.5u(k-1) + v_o(k)$$

Noise (colored, ARMA(1,2))

with $\mathbb{P} = [-1, 1]$ and

$$a_2(p_k) = \frac{1}{8}p_k^2$$

$$b_0(p_k) = \begin{cases} +0.5 & \text{if } p_k > +0.5 \\ p_k & \text{if } |p_k| \leq 0.5 \\ -0.5 & \text{if } p_k < -0.5 \end{cases}$$

Estimate the nonlinear dependencies under this noise!

Estimation under general noise conditions

Example: LPV-BJ identification

- Data records: white input, scheduling and noise signals with distributions $u(k) \in \mathcal{U}(-1,1)$, $p(k) \in \mathcal{U}(-1,1)$ and SNR=10dB.
- 100 Monte-Carlo runs:
  - Optimized RBF kernels K_2, K_3 (for the LS-SVM) with

$$\sigma_2 = 1.7, \ \sigma_3 = 0.5$$

  - Tuned regularization parameter

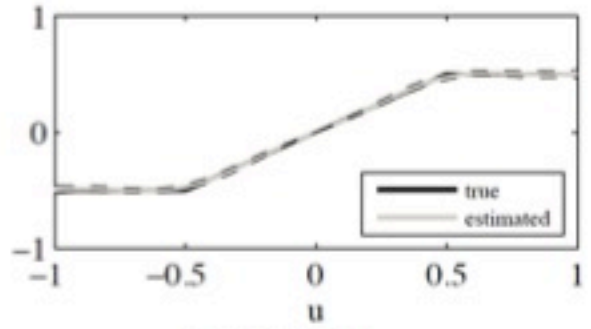
$$\frac{\gamma_{\mathrm{LS}}}{N} = 12.5, \quad \frac{\gamma_{\mathrm{IV}}}{N^2} = 16.5$$

  - 100 data records $\mathcal{D}_N$ with $N = 1000$ samples
  - Dynamic structure (model order) is assumed to be known

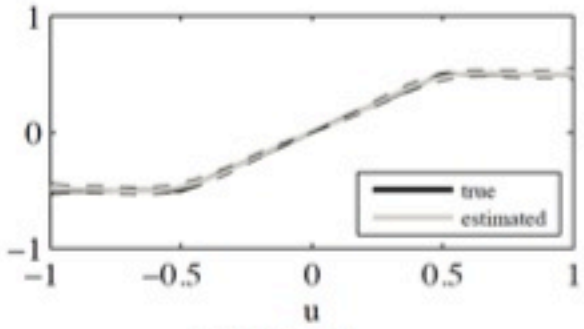
Estimation under general noise conditions

Example: LPV-BJ identification

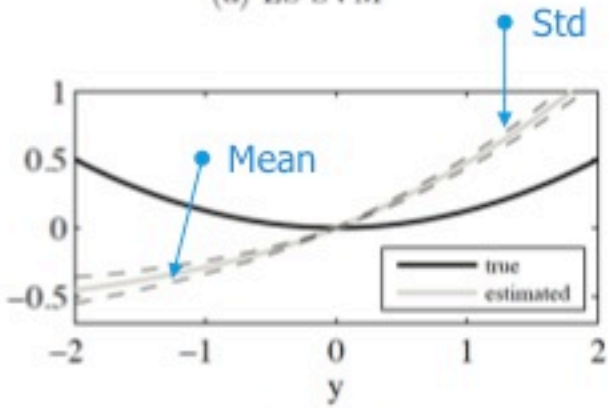
- Function estimates



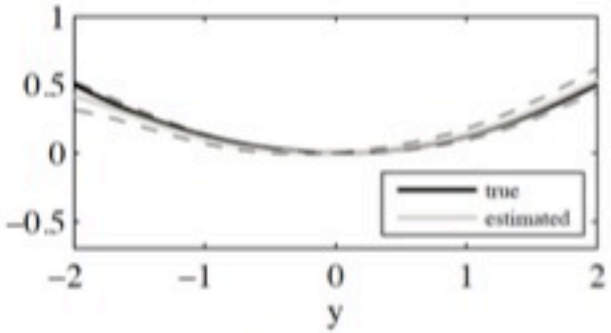
(a) LS-SVM



(b) IV-SVM



(c) LS-SVM



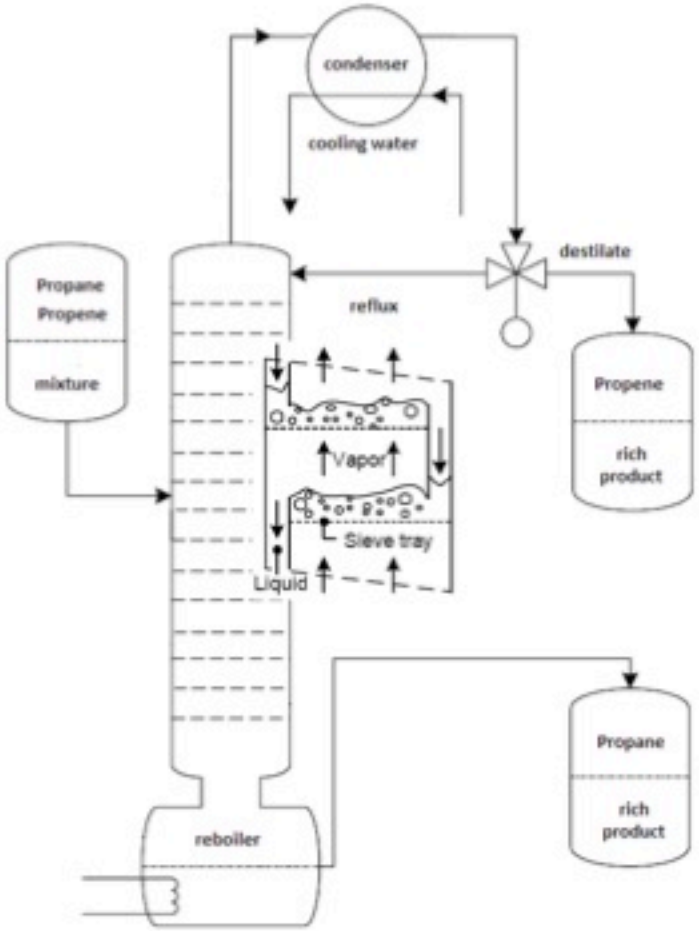
(d) IV-SVM

Estimation under general noise conditions

Example: High-purity distillation column

- 100 tray binary-column model (benchmark, ABB)
- High-purity distillation (PP splitter)
- Difficult nonlinear behavior at high-purity operating conditions





Estimation under general noise conditions

Example: High-purity distillation column

- LPV modeling setting
  - Inputs: the liquid flow rate and the vapor flow rate (LV-control)
  - Outputs: composition of top and bottom products (w.r.t. the most volatile component)
  - Scheduling: same as the outputs (purity levels)

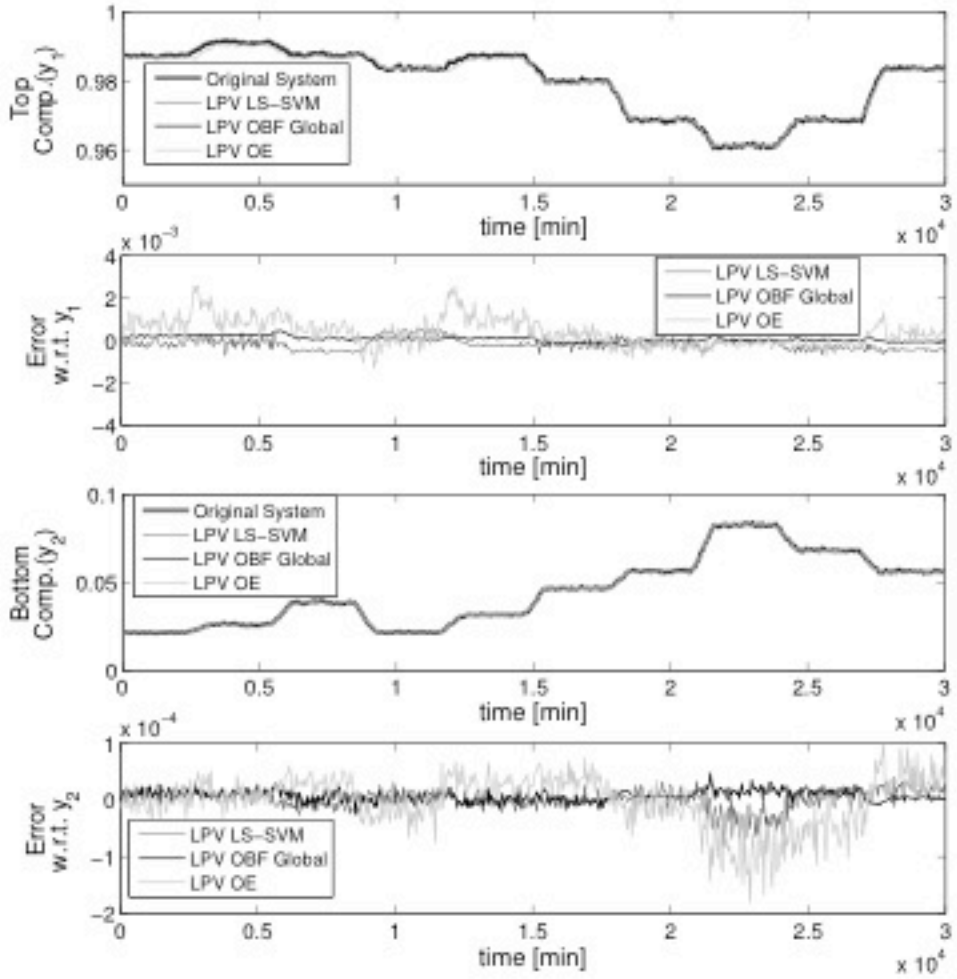
$$\mathbb{P} = [0.95,\ 0.995] \times [0.02,\ 0.1]$$

- Global identification:
  - 2500 data points, 15 min sampling time (8.7 days exp. campaign)
  - Low amplitude RBS input on an op profile with SNR $\approx$ 15 dB
  - Significant OE noise

Estimation under general noise conditions

Example: High-purity distillation column

- Results



Multidimensional kernels with inadequate excitation.
How to overcome this problem?

		LPV LS-SVM	LPV OE	global LPV-OBF
y_1	Avg. BFR	97.02%	92.49%	97.68%
	STD BFR	0.34%	0.23%	0.29%
	Avg. MSE	$8.27 \cdot 10^{-8}$	$5.03 \cdot 10^{-7}$	$5.14 \cdot 10^{-8}$
	STD MSE	$1.97 \cdot 10^{-8}$	$5.79 \cdot 10^{-8}$	$1.30 \cdot 10^{-8}$
y_2	Avg. BFR	99.89%	99.77%	99.90%
	STD BFR	0.02%	0.02%	0.03%
	Avg. MSE	$5.14 \cdot 10^{-10}$	$2.02 \cdot 10^{-9}$	$3.79 \cdot 10^{-10}$
	STD MSE	$2.02 \cdot 10^{-10}$	$4.79 \cdot 10^{-10}$	$2.43 \cdot 10^{-10}$

Contents

- The LPV system concept
- An LPV lest-squares SVM estimator
- Joint model order selection
- Estimation under general noise conditions
- Conclusions

Conclusions

- Model structure selection in LPV identification has been handled by LS-SVM approaches in terms of
  - Learning the coefficient dependencies
  - Learning the actual model order
- The provided non-parametric approaches are also applicable for general NL regression problems. LS-SVM's have been used in the NL identification context for quite some time…
- However, noise conditions restrict the learning capabilities of LS-SVM's as correlation between the noise and the kernel arguments can introduce bias.
- An IV-SVM has been introduced that is not affected by noise bias and has the same complexity as the original LS-SVM estimator.