A Bias-Correction Approach for the Identification of Piecewise Affine Output-Error Models

Manas Mejari[1] Valentina Breschi[2] Vihangkumar V. Naik[3]
Dario Piga[1]

[1]IDSIA Dalle Molle Institute for Artificial Intelligence, SUPSI-USI, Switzerland.

[2]Politecnico di Milano, Italy.

[3]ODYS S.r.l., Italy.

IFAC World Congress 2020, Berlin, Germany.

USI/SUPSI IDSIA Istituto Dalle Molle di studi sull'intelligenza artificiale

POLITECNICO MILANO

ODYS Advanced Controls & Optimization

Outline

Problem: PWA-OE identification

PWA-OE identification algorithm

- Stage S1: Bias-corrected least squares and iterative clustering
- Stage S2: Partitioning the regressor space

Numerical Example

Conclusions

Outline

Problem: PWA-OE identification

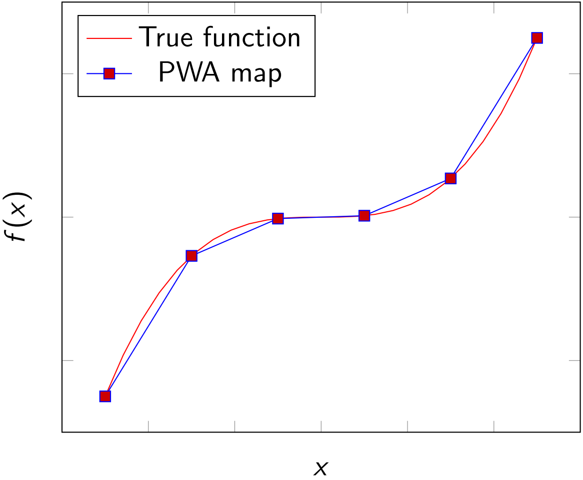
PWA-OE identification algorithm

- Stage S1: Bias-corrected least squares and iterative clustering
- Stage S2: Partitioning the regressor space

Numerical Example

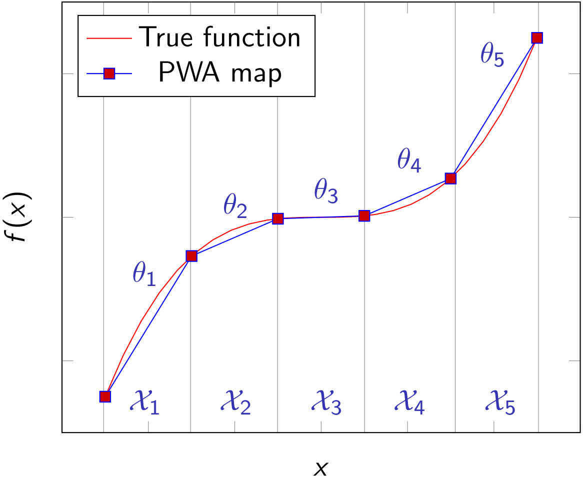
Conclusions

PWA models



- PWA maps have universal approximation property.
- Equivalence between PWA and hybrid models.

PWA models



$$f(x) = \begin{cases} \theta_1 \; x & \text{if } x \in \mathcal{X}_1 \\ \vdots & \\ \theta_s \; x & \text{if } x \in \mathcal{X}_s \end{cases}$$

Problem: PWA-OE identification

PWA Output-Error dynamical system:

$$y_{\mathrm{o}}(k) = f(x_{\mathrm{o}}(k))$$
$$y(k) = y_{\mathrm{o}}(k) + e_{\mathrm{o}}(k)$$

$$f(x_{\mathrm{o}}) = \begin{cases} (\theta_1^{\mathrm{o}})^\top \begin{bmatrix} x_{\mathrm{o}} \\ 1 \end{bmatrix} & \text{if } x_{\mathrm{o}} \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s^{\mathrm{o}})^\top \begin{bmatrix} x_{\mathrm{o}} \\ 1 \end{bmatrix} & \text{if } x_{\mathrm{o}} \in \mathcal{X}_s \end{cases}$$

$x_{\mathrm{o}}(k)$ is the noise-free regressor

$$x_{\mathrm{o}}(k) = [y_{\mathrm{o}}(k-1) \cdots y_{\mathrm{o}}(k-n_a)\ u(k-1) \cdots u(k-n_b)]^\top$$

$\{\mathcal{X}_i\}_{i=1}^s$ are the polyhedral partition of the regressor space.

Problem: PWA-OE identification

PWA Output-Error dynamical system:

$$y_{\mathrm{o}}(k) = f(x_{\mathrm{o}}(k))$$
$$y(k) = y_{\mathrm{o}}(k) + e_{\mathrm{o}}(k)$$

$$f(x_{\mathrm{o}}) = \begin{cases} (\theta_1^{\mathrm{o}})^\top \begin{bmatrix} x_{\mathrm{o}} \\ 1 \end{bmatrix} & \text{if } x_{\mathrm{o}} \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s^{\mathrm{o}})^\top \begin{bmatrix} x_{\mathrm{o}} \\ 1 \end{bmatrix} & \text{if } x_{\mathrm{o}} \in \mathcal{X}_s \end{cases}$$

$x_{\mathrm{o}}(k)$ is the noise-free regressor

$$x_{\mathrm{o}}(k) = [y_{\mathrm{o}}(k-1) \cdots y_{\mathrm{o}}(k-n_a) \ u(k-1) \cdots u(k-n_b)]^\top$$

$\{\mathcal{X}_i\}_{i=1}^s$ are the polyhedral partition of the regressor space.

Problem: PWA-OE identification

PWA Output-Error dynamical system:

$$\begin{aligned} y_{\rm o}(k) &= f(x_{\rm o}(k)) \\ y(k) &= y_{\rm o}(k) + e_{\rm o}(k) \end{aligned}$$

$$f(x_{\rm o}) = \begin{cases} (\theta_1^{\rm o})^\top \begin{bmatrix} x_{\rm o} \\ 1 \end{bmatrix} & \text{if } x_{\rm o} \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s^{\rm o})^\top \begin{bmatrix} x_{\rm o} \\ 1 \end{bmatrix} & \text{if } x_{\rm o} \in \mathcal{X}_s \end{cases}$$

$x_{\rm o}(k)$ is the noise-free regressor

$$x_{\rm o}(k) = [y_{\rm o}(k-1) \cdots y_{\rm o}(k-n_a) \; u(k-1) \cdots u(k-n_b)]^\top$$

$\{\mathcal{X}_i\}_{i=1}^s$ are the polyhedral partition of the regressor space.

Problem: PWA-OE identification

Model structure:

$$y(k) = \begin{cases} (\theta_1)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_s \end{cases}$$

$x(k)$ is the measured noisy regressor

$$x(k) = [y(k-1) \cdots y(k-n_a) \; u(k-1) \cdots u(k-n_b)]^\top$$

$\epsilon(k)$ is the residual, not necessarily white noise as in PWA-ARX.

Problem: PWA-OE identification

Model structure:

$$y(k) = \begin{cases} (\theta_1)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_s \end{cases}$$

$x(k)$ is the measured noisy regressor

$$x(k) = [y(k-1) \cdots y(k-n_a)\ u(k-1) \cdots u(k-n_b)]^\top$$

$\epsilon(k)$ is the residual, not necessarily white noise as in PWA-ARX.

Problem: PWA-OE identification

Model structure:

$$y(k) = \begin{cases} (\theta_1)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_s \end{cases}$$

$x(k)$ is the measured noisy regressor

$$x(k) = [y(k-1) \cdots y(k-n_a)\ u(k-1) \cdots u(k-n_b)]^\top$$

$\epsilon(k)$ is the residual, not necessarily white noise as in PWA-ARX.

Problem: PWA-OE identification

Model structure:

$$y(k) = \begin{cases} (\theta_1)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_1 \\ \vdots & \\ (\theta_s)^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k) & \text{if } x(k) \in \mathcal{X}_s \end{cases}$$

Goal:

Given $\{u(k), y(k)\}_{k=1}^N$ and the model structure n_a, n_b, s

- compute consistent estimates of $\{\theta_i^o\}_{i=1}^s$
- find a polyhedral partition $\{\mathcal{X}_i\}_{i=1}^s$.

Outline

Problem: PWA-OE identification

PWA-OE identification algorithm

- Stage S1: Bias-corrected least squares and iterative clustering
- Stage S2: Partitioning the regressor space

Numerical Example

Conclusions

Algorithm: PWA-OE identification

Stage S1.
- estimation of the bias-corrected model parameters $\{\theta_i\}_{i=1}^{s}$.
- simultaneous clustering of the regressors.

Stage S2.
- computation of polyhedral partition of the regressor space $\{\mathcal{X}_i\}_{i=1}^{s}$ using multi-category discrimination.

Algorithm: PWA-OE identification

Stage S1.
- estimation of the bias-corrected model parameters $\{\theta_i\}_{i=1}^{s}$.
- simultaneous clustering of the regressors.

Stage S2.
- computation of polyhedral partition of the regressor space $\{\mathcal{X}_i\}_{i=1}^{s}$ using multi-category discrimination.

Stage S1. Bias-corrected least squares estimates

Given the partition $\{\mathcal{X}_i\}_{i=1}^{s}$, LS estimate of the i-th model is

$$\theta_i^{\mathrm{LS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}$$

$\mathbb{X}_i, \mathbb{Y}_i$ measured noisy regressors, output of the i-th model.

- It can be proved that LS estimate θ_i^{LS} is biased!
- Asymptotic bias is caused due to OE structure:

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} = \theta_i^{\mathrm{o}} + \lim_{N_i \to \infty} \underbrace{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\mathrm{bias}}$$

Stage S1. Bias-corrected least squares estimates

Given the partition $\{\mathcal{X}_i\}_{i=1}^{s}$, LS estimate of the i-th model is

$$\theta_i^{\mathrm{LS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}$$

$\mathbb{X}_i, \mathbb{Y}_i$ measured noisy regressors, output of the i-th model.

- It can be proved that LS estimate θ_i^{LS} is biased!
- Asymptotic bias is caused due to OE structure:

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} = \theta_i^{\mathrm{o}} + \lim_{N_i \to \infty} \underbrace{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\mathrm{bias}}$$

Stage S1. Bias-corrected least squares estimates

Given the partition $\{\mathcal{X}_i\}_{i=1}^{s}$, LS estimate of the i-th model is

$$\theta_i^{\mathrm{LS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}$$

$\mathbb{X}_i, \mathbb{Y}_i$ measured noisy regressors, output of the i-th model.

- It can be proved that LS estimate θ_i^{LS} is biased!
- Asymptotic bias is caused due to OE structure:

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} = \theta_i^{\mathrm{o}} + \lim_{N_i \to \infty} \underbrace{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\text{bias}}$$

Stage S1. Bias-corrected least squares estimates

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} = \theta_i^{\mathrm{o}} + \lim_{N_i \to \infty} \underbrace{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\mathrm{bias}}$$

Asymptotic bias $B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})$ does not to converge to 0.

- Thus, the LS estimates are not consistent

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} \neq \theta_i^{\mathrm{o}},$$

even for a true partition!

Stage S1. Bias-corrected least squares estimates

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} = \theta_i^{\mathrm{o}} + \lim_{N_i \to \infty} \underbrace{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\text{bias}}$$

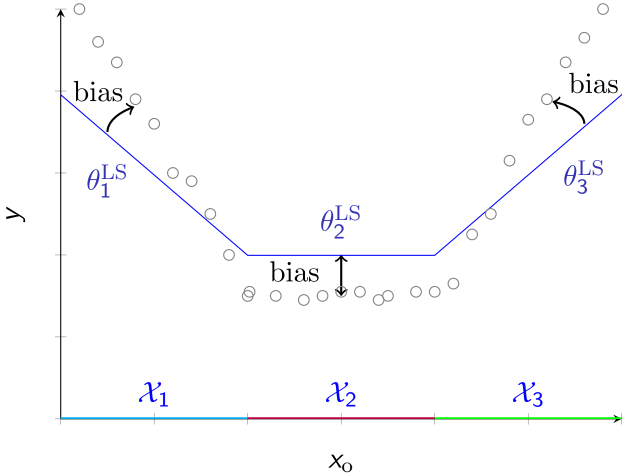
Asymptotic bias $B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})$ does not to converge to 0.

- Thus, the LS estimates are not consistent

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} \neq \theta_i^{\mathrm{o}},$$

even for a true partition!

Stage S1. Bias-corrected least squares estimates



Stage S1. Bias-corrected least squares estimates

Bias-correction: Quantify and remove the bias from LS estimates

$$\theta_i^{\mathrm{o}} = \theta_i^{\mathrm{LS}} - \underbrace{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\text{bias}}$$

However, the bias $B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})$ depends on θ_i^{o} and can not be computed.

- Key idea: We define the corrected LS estimate θ_i^{CLS}

$$\theta_i^{\mathrm{CLS}} = \theta_i^{\mathrm{LS}} - B_\Delta(\theta_i^{\mathrm{CLS}}, \mathbb{X}_{i,\mathrm{o}})$$

Stage S1. Bias-corrected least squares estimates

Bias-correction: Quantify and remove the bias from LS estimates

$$\theta_i^{\mathrm{o}} = \theta_i^{\mathrm{LS}} - \underbrace{B_{\Delta}(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\text{bias}}$$

However, the bias $B_{\Delta}(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})$ depends on θ_i^{o} and can not be computed.

- Key idea: We define the corrected LS estimate θ_i^{CLS}

$$\theta_i^{\mathrm{CLS}} = \theta_i^{\mathrm{LS}} - B_{\Delta}(\theta_i^{\mathrm{CLS}}, \mathbb{X}_{i,\mathrm{o}})$$

Stage S1. Bias-corrected least squares estimates

Bias-correction: Quantify and remove the bias from LS estimates

$$\theta_i^{\mathrm{o}} = \theta_i^{\mathrm{LS}} - \underbrace{B_{\Delta}(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})}_{\mathrm{bias}}$$

However, the bias $B_{\Delta}(\theta_i^{\mathrm{o}}, \mathbb{X}_{i,\mathrm{o}})$ depends on θ_i^{o} and can not be computed.

- Key idea: We define the corrected LS estimate θ_i^{CLS}

$$\theta_i^{\mathrm{CLS}} = \theta_i^{\mathrm{LS}} - B_{\Delta}(\theta_i^{\mathrm{CLS}}, \mathbb{X}_{i,\mathrm{o}})$$

Stage S1. Bias-corrected least squares estimates

The corrected LS estimate θ_i^{CLS}

$$\begin{aligned}\theta_i^{\mathrm{CLS}} &= \theta_i^{\mathrm{LS}} - B_\Delta(\theta_i^{\mathrm{CLS}}, \mathbb{X}_{i,\mathrm{o}}) \\ &= \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \mathbb{X}_i^\top \Delta\mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.\end{aligned}$$

where $\Delta\mathbb{X}_i = \mathbb{X}_{i,\mathrm{o}} - \mathbb{X}_i$.

However, θ_i^{CLS} still can not be computed as it depends on noise-free regressors $\mathbb{X}_{i,\mathrm{o}}$.

Stage S1. Bias-corrected least squares estimates

The corrected LS estimate θ_i^{CLS}

$$\begin{aligned}
\theta_i^{\mathrm{CLS}} &= \theta_i^{\mathrm{LS}} - B_\Delta(\theta_i^{\mathrm{CLS}}, \mathbb{X}_{i,\mathrm{o}}) \\
&= \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \mathbb{X}_i^\top \Delta \mathbb{X}_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.
\end{aligned}$$

where $\Delta\mathbb{X}_i = \mathbb{X}_{i,\mathrm{o}} - \mathbb{X}_i$.

However, θ_i^{CLS} still can not be computed as it depends on noise-free regressors $\mathbb{X}_{i,\mathrm{o}}$.

Stage S1. Bias-corrected least squares estimates

The corrected LS estimate θ_i^{CLS}

$$\theta_i^{\mathrm{CLS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \mathbb{X}_i^\top \Delta \mathbb{X}_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.$$

Key idea: We replace the term $\mathbb{X}_i^\top \Delta \mathbb{X}_i$ with a bias-eliminating matrix Ψ_i

$$\lim_{N_i \to \infty} \frac{1}{N_i} \mathbb{X}_i^\top \Delta \mathbb{X}_i = \lim_{N_i \to \infty} \frac{1}{N_i} \Psi_i \quad \text{w.p. 1.}$$

- Ψi can be computed from the available information.
- In our approach, Ψi depends on the noise-variance, which is assumed to be known.

Stage S1. Bias-corrected least squares estimates

The corrected LS estimate θ_i^{CLS}

$$\theta_i^{\mathrm{CLS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \mathbb{X}_i^\top \Delta \mathbb{X}_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.$$

Key idea: We replace the term $\mathbb{X}_i^\top \Delta \mathbb{X}_i$ with a bias-eliminating matrix Ψ_i

$$\lim_{N_i \to \infty} \frac{1}{N_i} \mathbb{X}_i^\top \Delta \mathbb{X}_i = \lim_{N_i \to \infty} \frac{1}{N_i} \Psi_i \text{ w.p. 1.}$$

- Ψi can be computed from the available information.
- In our approach, Ψi depends on the noise-variance, which is assumed to be known.

Stage S1. Bias-corrected least squares estimates

The corrected LS estimate θ_i^{CLS}

$$\theta_i^{\mathrm{CLS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \mathbb{X}_i^\top \Delta \mathbb{X}_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.$$

Key idea: We replace the term $\mathbb{X}_i^\top \Delta \mathbb{X}_i$ with a bias-eliminating matrix Ψ_i

$$\lim_{N_i \to \infty} \frac{1}{N_i} \mathbb{X}_i^\top \Delta \mathbb{X}_i = \lim_{N_i \to \infty} \frac{1}{N_i} \Psi_i \ \text{ w.p. 1.}$$

- Ψi can be computed from the available information.
- In our approach, Ψi depends on the noise-variance, which is assumed to be known.

Stage S1. Bias-corrected least squares estimates

The bias-corrected estimate θ_i^{BC}

$$\theta_i^{\mathrm{BC}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \Psi_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.$$

with $\Psi i = -\sigma_e^2 N_i \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$.

The bias-corrected estimate θ_i^{BC} is consistent!

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{BC}} = \theta_i^{\mathrm{o}}, \quad \text{w.p. } 1$$

Stage S1. Bias-corrected least squares estimates

The bias-corrected estimate θ_i^{BC}

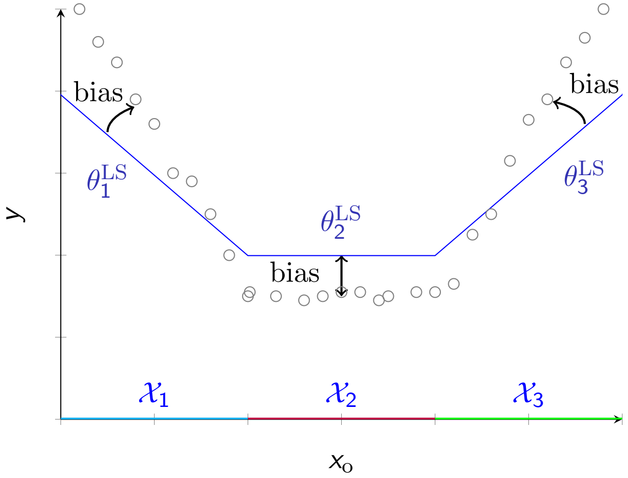
$$\theta_i^{\mathrm{BC}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \Psi_i}{N_i} \right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}.$$

with $\Psi i = -\sigma_e^2 N_i \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$.

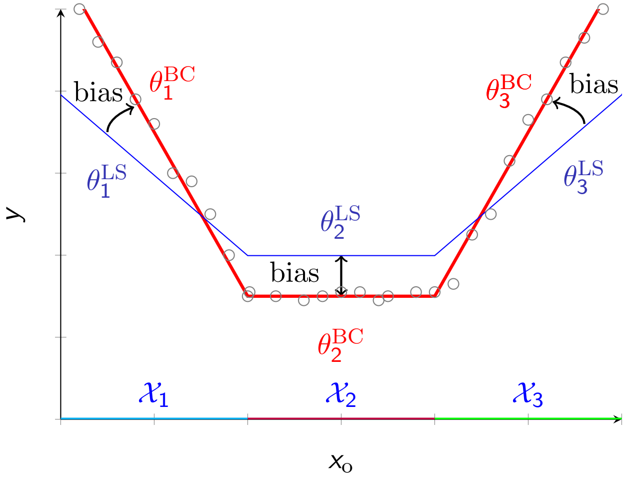
The bias-corrected estimate θ_i^{BC} is consistent!

$$\lim_{N_i \to \infty} \theta_i^{\mathrm{BC}} = \theta_i^{\mathrm{o}}, \quad \text{w.p. } 1$$

Stage S1. Bias-corrected least squares estimates



Stage S1. Bias-corrected least squares estimates



Stage S1. Clustering the regressors

The active mode $\sigma(k)$ is defined as

$$\sigma(k) = i \quad \Leftrightarrow \quad x_{\mathrm{o}}(k) \in \mathcal{X}_i,$$

The proposed clustering criterion to estimate $\sigma(k)$

$$\underbrace{\sigma(k)}_{\text{active mode}} \leftarrow \arg\min_{i=1,\ldots,s} \lambda e_i^2(k) + \|\hat{x}(k) - c_i\|_2^2;$$

- $e_i(k) = y(k) - (\theta_i^{\mathrm{BC}})^\top \begin{bmatrix} \hat{x}(k) \\ 1 \end{bmatrix}$: prediction error.
- $\|\hat{x}(k) - c_i\|_2^2$: distance from cluster's centroid c_i

Stage S1. Clustering the regressors

The active mode $\sigma(k)$ is defined as

$$\sigma(k) = i \quad \Leftrightarrow \quad x_{\mathrm{o}}(k) \in \mathcal{X}_i,$$

The proposed clustering criterion to estimate $\sigma(k)$

$$\underbrace{\sigma(k)}_{\text{active mode}} \quad \leftarrow \arg \min_{i=1,\ldots,s} \; \lambda e_i^2(k) + \|\hat{x}(k) - c_i\|_2^2;$$

- $e_i(k) = y(k) - (\theta_i^{\mathrm{BC}})^\top \begin{bmatrix} \hat{x}(k) \\ 1 \end{bmatrix}$: prediction error.
- $\|\hat{x}(k) - c_i\|_2^2$: distance from cluster's centroid c_i

Stage S1. Clustering the regressors

The active mode $\sigma(k)$ is defined as

$$\sigma(k) = i \quad \Leftrightarrow \quad x_{\mathrm{o}}(k) \in \mathcal{X}_i,$$

The proposed clustering criterion to estimate $\sigma(k)$

$$\underbrace{\sigma(k)}_{\text{active mode}} \leftarrow \arg \min_{i=1,\dots,s} \lambda e_i^2(k) + \|\hat{x}(k) - c_i\|_2^2;$$

- $e_i(k) = y(k) - (\theta_i^{\mathrm{BC}})^\top \begin{bmatrix} \hat{x}(k) \\ 1 \end{bmatrix}$: prediction error.
- $\|\hat{x}(k) - c_i\|_2^2$: distance from cluster's centroid c_i

Stage S1. Iterative parameter estimation and clustering

1. **iterate for** $m = 1, \dots, M$ **do**

   1.1 **estimate** $\{\theta_i^m\}_{i=1}^s$ for fixed $\{\sigma^{m-1}(k)\}_{k=1}^N$:

   $$\theta_i^m = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \Psi_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i};$$

   1.2 **estimate** $\{\sigma^m(k)\}_{k=1}^N$ for a given $\{\theta_i^m\}_{i=1}^s$:

   1.2.1 **for** $k = 1, \dots, N$ **do**
   $\sigma(k) \leftarrow \arg\min_{i=1,\dots,s} \ \lambda e_i^2(k) + \|\hat{x}(k) - c_i\|_2^2$
   update centroids c_i, $\hat{x}(k)$.

2. **end for**;

Summary of Stage S1

- The bias-corrected parameters θ_i of PWA-OE model are estimated.
- The estimated regressors $\{\hat{x}(k)\}_{k=1}^{N}$ are clustered into s clusters based on $\sigma(k)$.
- Each cluster corresponds to a polyhedral partition $\mathcal{X}_i$.
- Next stage: Compute polyhedral partition $\mathcal{X}_i$ using linear multicategory discrimination

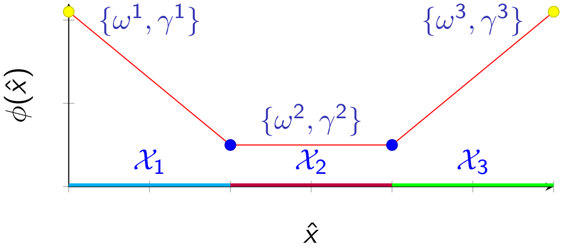
Summary of Stage S1

- The bias-corrected parameters θ_i of PWA-OE model are estimated.
- The estimated regressors $\{\hat{x}(k)\}_{k=1}^{N}$ are clustered into s clusters based on $\sigma(k)$.
- Each cluster corresponds to a polyhedral partition $\mathcal{X}_i$.
- Next stage: Compute polyhedral partition $\mathcal{X}_i$ using linear multicategory discrimination

Stage S2: Linear multicategory discrimination

Compute polyhedral partition of space $\hat{\mathcal{X}}$ characterized by the PWA separator function ϕ

$$\phi(\hat{x}) = \max_{i=1,\ldots,s} \left(\begin{bmatrix} \hat{x}^\top & -1 \end{bmatrix} \begin{bmatrix} \omega^i \\ \gamma^i \end{bmatrix} \right),$$



Stage S2: Linear multicategory discrimination

The parameters $\{\omega^i, \gamma^i\}_{i=1}^{s}$ are computed by solving the optimization problem (Breschi et al., Automatica, 2016)

$$\min_{\omega^i, \gamma^i} \frac{\kappa}{2} \sum_{i=1}^{s} \left(\|\omega^i\|_2^2 + (\gamma^i)^2 \right) + \sum_{i=1}^{s} \sum_{\substack{j=1 \\ j \neq i}}^{s} \frac{1}{N_i} \left\| \left(\begin{bmatrix} M_i & -1_{N_i} \end{bmatrix} \begin{bmatrix} \omega^j - \omega^i \\ \gamma^j - \gamma^i \end{bmatrix} + 1_{N_i} \right)_+ \right\|_2^2$$

Violation of inequalities defining the polytopes are penalized.
$\kappa > 0$ ensures the optimization problem is strongly convex.

Stage S2: Linear multicategory discrimination

The parameters $\{\omega^i, \gamma^i\}_{i=1}^s$ are computed by solving the optimization problem (Breschi et al., Automatica, 2016)

$$\min_{\omega^i, \gamma^i} \frac{\kappa}{2} \sum_{i=1}^{s} \left(\|\omega^i\|_2^2 + (\gamma^i)^2\right) + \sum_{i=1}^{s} \sum_{\substack{j=1 \\ j \neq i}}^{s} \frac{1}{N_i} \left\| \left(\begin{bmatrix} M_i & -1_{N_i} \end{bmatrix} \begin{bmatrix} \omega^j - \omega^i \\ \gamma^j - \gamma^i \end{bmatrix} + 1_{N_i} \right)_+ \right\|_2^2$$

Violation of inequalities defining the polytopes are penalized.

$\kappa > 0$ ensures the optimization problem is strongly convex.

Stage S2: Linear multicategory discrimination

The parameters $\{\omega^i, \gamma^i\}_{i=1}^s$ are computed by solving the optimization problem (Breschi et al., Automatica, 2016)

$$\min_{\omega^i, \gamma^i} \frac{\kappa}{2} \sum_{i=1}^{s} \left(\|\omega^i\|_2^2 + (\gamma^i)^2 \right) +$$

$$\sum_{i=1}^{s} \sum_{\substack{j=1 \\ j \neq i}}^{s} \frac{1}{N_i} \left\| \left(\left[\, M_i \;\; -1_{N_i} \,\right] \left[\begin{smallmatrix} \omega^j - \omega^i \\ \gamma^j - \gamma^i \end{smallmatrix} \right] + 1_{N_i} \right)_+ \right\|_2^2$$

Violation of inequalities defining the polytopes are penalized.
$\kappa > 0$ ensures the optimization problem is strongly convex.

Outline

Numerical Example

PWA-OE data-generating system

$$y_o(k) = \begin{cases} \begin{bmatrix} -0.4 & 1 & 1.5 \end{bmatrix} x_o(k), & \text{if } 4y_o(k-1) - u(k-1) + 10 < 0, \\ \begin{bmatrix} 0.5 & -1 & -1.5 \end{bmatrix} x_o(k), & \text{if } 4y_o(k-1) - u(k-1) + 10 < 0, \\ & \&5y_o(k-1) + u(k-1) - 6 \leq 0, \\ \begin{bmatrix} -0.3 & 0.5 & -1.7 \end{bmatrix} x_o(k), & \text{if } 5y_o(k-1) + u(k-1) - 6 > 0, \end{cases}$$

$$y(k) = y_o(k) + e_o(k)$$

with $N = 5000$ training samples and SNR $= 11.7$ dB.

- Identification of PWA model with $s = 3$, $n_a = n_b = 1$
- Stage S1 algorithm is run for 20 iterations, with $\lambda = \sigma_e^{-2} = 1.56$ and random initial guess $\{\sigma^0(k)\}_{k=1}^{N}$.
- Stage S2 is executed with $\kappa = 10^{-5}$.

Numerical Example

PWA-OE data-generating system

$$y_{\mathrm{o}}(k) = \begin{cases} \begin{bmatrix} -0.4 & 1 & 1.5 \end{bmatrix} x_{\mathrm{o}}(k), & \text{if } 4y_{\mathrm{o}}(k-1) - u(k-1) + 10 < 0, \\ \begin{bmatrix} 0.5 & -1 & -1.5 \end{bmatrix} x_{\mathrm{o}}(k), & \text{if } 4y_{\mathrm{o}}(k-1) - u(k-1) + 10 < 0, \\ & \&5y_{\mathrm{o}}(k-1) + u(k-1) - 6 \leq 0, \\ \begin{bmatrix} -0.3 & 0.5 & -1.7 \end{bmatrix} x_{\mathrm{o}}(k), & \text{if } 5y_{\mathrm{o}}(k-1) + u(k-1) - 6 > 0, \end{cases}$$

$$y(k) = y_{\mathrm{o}}(k) + e_{\mathrm{o}}(k)$$

with $N = 5000$ training samples and SNR $= 11.7$ dB.

- Identification of PWA model with $s = 3$, $n_a = n_b = 1$
- Stage S1 algorithm is run for 20 iterations, with $\lambda = \sigma_e^{-2} = 1.56$ and random initial guess $\{\sigma^0(k)\}_{k=1}^{N}$.
- Stage S2 is executed with $\kappa = 10^{-5}$.

Numerical Example

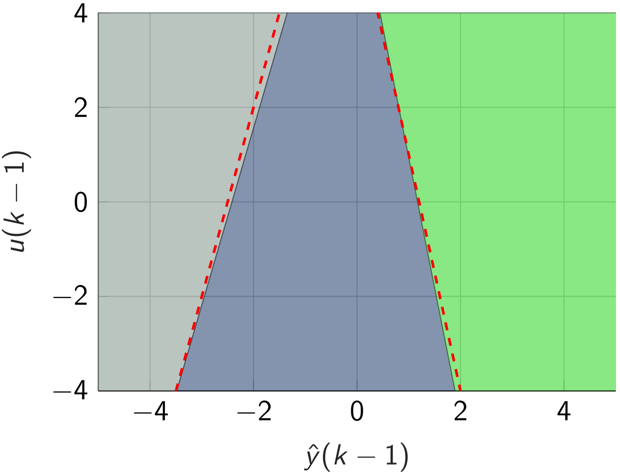
Norm of the parameter estimation error LS *vs* BC estimates.

Mode	$\left\|\theta^{\mathrm{o}} - \theta^{\mathrm{BC}}\right\|$	$\left\|\theta^{\mathrm{o}} - \theta^{\mathrm{LS}}\right\|$
$s = 1$	0.0196	0.7818
$s = 2$	0.0070	0.2091
$s = 3$	0.0275	0.2156

Mode fit index $\mathrm{MF} = \left(\frac{1}{N}\sum_{k=1}^{N}\mathbb{I}(\sigma(k) = \sigma^{\star}(k))\right) \times 100\%$

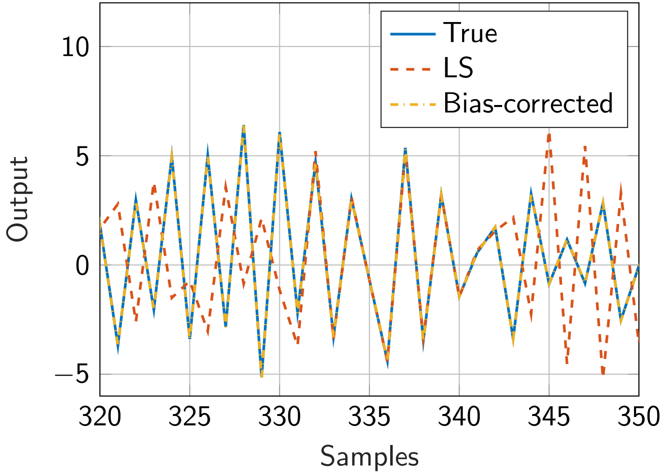
	BC	LS
Training	98.76%	98.50%
Validation	88.40%	77.80%

Numerical Example



True (dashed red lines) *vs* estimated polyhedral partition.

Numerical Example



Validation dataset: true *vs* simulated output of the LS model BC model.

Outline

Conclusions

- An iterative batch algorithm for identification of PWA-OE models is proposed.
- Bias-correction scheme is combined with a clustering algorithm for parameter estimation and regressors' clustering.
- The parameters are consistent under suitable assumptions.
- A partition of the regressor space is estimated with high accuracy by employing multicategory discrimination.
- Future work will be focused on data-driven model structure selection.

Conclusions

- An iterative batch algorithm for identification of PWA-OE models is proposed.
- Bias-correction scheme is combined with a clustering algorithm for parameter estimation and regressors' clustering.
- The parameters are consistent under suitable assumptions.
- A partition of the regressor space is estimated with high accuracy by employing multicategory discrimination.
- Future work will be focused on data-driven model structure selection.

Conclusions

- An iterative batch algorithm for identification of PWA-OE models is proposed.
- Bias-correction scheme is combined with a clustering algorithm for parameter estimation and regressors' clustering.
- The parameters are consistent under suitable assumptions.
- A partition of the regressor space is estimated with high accuracy by employing multicategory discrimination.
- Future work will be focused on data-driven model structure selection.

Conclusions

- An iterative batch algorithm for identification of PWA-OE models is proposed.
- Bias-correction scheme is combined with a clustering algorithm for parameter estimation and regressors' clustering.
- The parameters are consistent under suitable assumptions.
- A partition of the regressor space is estimated with high accuracy by employing multicategory discrimination.
- Future work will be focused on data-driven model structure selection.

Conclusions

- An iterative batch algorithm for identification of PWA-OE models is proposed.
- Bias-correction scheme is combined with a clustering algorithm for parameter estimation and regressors' clustering.
- The parameters are consistent under suitable assumptions.
- A partition of the regressor space is estimated with high accuracy by employing multicategory discrimination.
- Future work will be focused on data-driven model structure selection.

Thank You